



Zadání bakalářské práce

Název:	Vizualizace statistik českých hokejistů v NHL
Student:	Adam Lesch
Vedoucí:	Ing. Karel Klouda, Ph.D.
Studijní program:	Informatika
Obor / specializace:	Znalostní inženýrství
Katedra:	Katedra aplikované matematiky
Platnost zadání:	do konce letního semestru 2025/2026

Pokyny pro vypracování

Cílem práce je vytvoření webové prezentace dat (především) českých hráčů NHL. Výsledkem by měly být zejména vizualizace srozumitelně prezentující pokročilejší statistiky a pohledy na hráče než jakými jsou obvykle prezentované tabulky s počtem gólů, asistencí a pod.

- 1) Provedte rešerši webových stránek, které se zabývají vizualizací hokejových statistik (zejména těch pokročilých).
- 2) Provedte rešerši dostupných datových zdrojů (zejména dat poskytovaných přímo NHL) a prozkoumejte jejich vhodnost pro účel práce.
- 3) Pomocí vhodně zvolených nástrojů vytvořte sadu automaticky aktualizovaných vizualizací prezentujících výkony českých hráčů v NHL. Zaměřte se zejména na následující aspekty:
 - a) Jak se českým hráčům dařilo v jednotlivých zápasech.
 - b) Jak se daří a dařilo jednotlivým hráčům v dané sezóně.
 - c) Jak se v posledních letech vyvíjí role a pozice českých hokejistů v NHL jako celku.
- 4) Vytvořte webovou aplikaci prezentující vybrané vizualizace.

Bakalářská práce

VIZUALIZACE STATISTIK ČESKÝCH HOKEJISTŮ V NHL

Adam Lesch

Fakulta informačních technologií
Katedra aplikované matematiky
Vedoucí: Ing. Karel Klouda, Ph.D.
15. května 2024

České vysoké učení technické v Praze
Fakulta informačních technologií

© 2024 Adam Lesch. Všechna práva vyhrazena.

Tato práce vznikla jako školní dílo na Českém vysokém učení technickém v Praze, Fakultě informačních technologií. Práce je chráněna právními předpisy a mezinárodními úmluvami o právu autorském a právech souvisejících s právem autorským. K jejímu užití, s výjimkou bezúplatných zákonných licencí a nad rámec oprávnění uvedených v Prohlášení, je nezbytný souhlas autora.

Odkaz na tuto práci: Lesch Adam. *Vizualizace statistik českých hokejistů v NHL*. Bakalářská práce. České vysoké učení technické v Praze, Fakulta informačních technologií, 2024.

Obsah

Poděkování	v
Prohlášení	vi
Abstrakt	vii
Seznam zkratk	viii
Úvod	1
1 Lední hokej	2
1.1 Historie	2
1.2 NHL	2
1.3 Pravidla hry v NHL	3
2 Analýza webových zdrojů	4
2.1 Oficiální stránky NHL	4
2.1.1 NHL EDGE	4
2.2 Ostatní	5
2.2.1 NHL.CZ	5
2.2.2 Shiftchart	5
2.2.3 Hockeyviz	5
2.2.4 Interaktivní mapa	6
3 Analýza datových zdrojů	7
3.1 Pořadí týmů	7
3.2 Seznam sezón	7
3.3 Plán zápasů	8
3.4 Soupisky	9
3.5 Hráči	9
3.6 Střídání	9
3.7 Play-by-play	10
3.8 Zápasové záznamy	11
4 Použité nástroje	12
4.1 Hardware	12
4.2 Software	12
4.2.1 RStudio	13
4.2.2 MongoDB	14
4.2.3 Docker	14

5	Zpracování dat	15
5.1	Stahování dat	15
5.1.1	Čeští hráči	15
5.1.2	Zápasové akce	16
5.1.3	Střídání	17
5.1.4	Krátkodobé statistiky	17
5.1.5	Dlouhodobé statistiky	18
5.2	Tvorba datasetů	19
5.2.1	Dataset českých hráčů	19
5.2.2	Dataset „play-by-play“	20
5.2.3	Dataset času na ledě	21
5.2.4	Dataset dlouhodobých statistik	22
6	Vizualizace	24
6.1	Dnešní dění	24
6.1.1	Schéma střel	24
6.1.2	Schéma akcí	26
6.1.3	Čas na ledě	27
6.2	Krátkodobé trendy	27
6.3	Dlouhodobé trendy	29
	Závěr	31
	Obsah příloženého média	35

Seznam obrázků

2.1	Ukázka z webu <i>shiftchart.com</i> [16]	5
5.1	Ukázka formátu uložených informací o hráčích	16
6.1	Přehled střel ze dne 8. 2. 2024, zvolený hráč Martin Nečas	25
6.2	Přehled akcí ze dne 8. 2. 2024, zvolený hráč Jakub Lauko	26
6.3	Graf „Čas na ledě“ ze dne 8. 2. 2024, zvolený hráč Ondřej Palát	27
6.4	„Krátkodobé trendy“ pro bruslaře, zvolenou statistikou je <i>Čas na ledě</i> a hráčem Radek Faksa.	29
6.5	„Krátkodobé trendy“ pro brankáře, zvolenou statistikou je <i>Úspěšnost zákroků</i> , zvoleným hráčem Lukáš Dostál. Modrou barvou jsou vyznačené zápasy, ve kterých nestartoval.	29
6.6	„Dlouhodobé trendy“, zvolenou statistikou je <i>Počet hráčů</i>	30

Seznam tabulek

5.1	Označení identifikátorů hráčů v „play-by-play“ datech	17
5.2	Typy akcí	21
5.3	Typy střel	21

Seznam výpisů kódu

1	Ukázka záznamu o týmu	72	Ukázka záznamu plánu zápasů	83	Ukázka záznamu o střídáních	94	Ukázka záznamu o zápasových akcích	105	Ukázka zápasových záznamů	111
---	-----------------------	----	-----------------------------	----	-----------------------------	----	------------------------------------	-----	---------------------------	-----

Tímto bych rád poděkoval vedoucímu mé práce, Ing. Karlu Kloudovi, Ph.D., za ochotnou pomoc, vstřícnou komunikaci a věnovaný volný čas při pravidelných konzultacích, stejně jako poskytnutí cenných rad při počátečním výběru tématu práce. Dále bych chtěl poděkovat rodičům za neustálou podporu, nejen během vypracování bakalářské práce, ale během celého studia.

Prohlášení

Prohlašuji, že jsem předloženou práci vypracoval samostatně a že jsem uvedl veškeré použité informační zdroje v souladu s Metodickým pokynem o dodržování etických principů při přípravě vysokoškolských závěrečných prací.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona, ve znění pozdějších předpisů, zejména skutečnost, že České vysoké učení technické v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 citovaného zákona.

V Praze dne 15. května 2024

Abstrakt

Tato bakalářská práce se zabývá vizualizací hokejových statistik zejména českých hráčů ve formě webové aplikace. Nejprve byla provedena rešerše zaměřující se na existující webové stránky prezentující vizualizace hokejových statistik v NHL. Následně byly prozkoumány zdroje dat pro tvorbu vizualizací. Na základě rešerše byla vytvořena webová aplikace prezentující výkony českých hokejistů v NHL a nabízející přehled tvořený automaticky aktualizovanými grafy. Grafy se zaměřují na tři pohledy: výkony hráčů v jednotlivých zápasech, přehled hráčských statistik během aktuální sezóny a vývoj role českých hráčů jako celku v NHL. Aplikace byla vyvinuta v jazyce R, pomocí balíčku Shiny.

Klíčová slova vizualizační nástroje, sportovní statistiky, webová aplikace, hokej, NHL, čeští hokejisté, R, Shiny

Abstract

This bachelor thesis deals with visualization of hockey statistics with emphasis on Czech players in the form of a web application. First, research was conducted focusing on existing web pages presenting visualizations of hockey statistics in the NHL. Subsequently, data sources for creating visualizations were explored. The primary analysis informed the development of a web application, which presents performances of Czech hockey players in the NHL and offers an overview consisting of automatically updated plots. The plots explore three aspects: the performance of players in individual games, an overview of player statistics during the current season and the evolution of the role of Czech players as a whole in the NHL. The application was developed in R, using the Shiny package.

Keywords visualization tools, sport statistics, web application, hockey, NHL, czech players, R, Shiny

Seznam zkratek

API	Application Programming Interface
BSON	Binary JSON
CRAN	The Comprehensive R Archive Network
FTP	File Transfer Protocol
GNU	GNU's Not Unix
HTTP	Hypertext Transfer Protocol
IDE	Integrated Development Environment
IIHF	International Ice Hockey Federation
JSON	JavaScript Object Notation
MLB	Major League Baseball
NBA	National Basketball Association
NFL	National Football Association
NHA	National Hockey Association
NHL	National Hockey League
NoSQL	not only SQL
URL	Uniform Resource Locator
WHA	World Hockey Association

Úvod

Publikací knihy *Bill James Historical Abstract* [1] v roce 1985, byla širší veřejnost poprvé představena novému přístupu ke statistické analýze ve sportu. Autor Bill James, baseballový novinář, historik a statistik, se ve své knize zabývá především netradičními otázkami ohledně baseballové taktiky, na které hledá matematicky podložené odpovědi. James byl „*průkopník ve světě statistik a měřítek, které nabízejí unikátní vzhled do výkonu hráčů*“ [2]. V knize je mimo jiné popsána statistika *Win Shares* (Podíl na vítězství). Jedná se o Jamesův vlastní vynález, jehož účelem je vyjádřit jedním číslem vliv každého z hráčů na výsledek zápasu [3]. Tato statistika pak sloužila jako jedno z měřítek, které James následně využil v další části své knihy, zabývající se žebříčky nejúspěšnějších hráčů na každé pozici.

Z původních myšlenek Jamese vychází o několik let později Billy Beane, manažer baseballového týmu Oakland Athletics. Příběh tohoto týmu během sezóny 2002 vypráví kniha *Moneyball*, od autora Michaela Lewise [4]. Přestože se v té době jednalo o jeden z nechudších týmů v lize, co se týče rozpočtu na platy hráčů, dosahoval velmi dobrých výsledků. Hlavní podíl na tom má právě Beane, jehož inovativní přístup se zaměřil na „odpadlíky“ – hráče, které ostatní týmy opomíjely, ačkoliv pokročilé statistiky vypovídaly o jejich pozitivním přínosu. [5]

„Napodobení je neupřímnější formou obdivu.“ Ve vrcholovém sportu to platí dvojnásob a zanedlouho procházejí i ostatní sporty svou tzv. „datovou revolucí.“ Nový trend je nejzřetelnější v prostředí severoamerických sportovních lig. Konkrétně v NHL to znamená, že každý z 32 zimních stadionů je vybaven až dvaceti kamerami snímající infračervené záření z elektronických čipů v puku a všitých do dresů jednotlivých hráčů [6, 7]. Třicetkrát za vteřinu jsou tímto způsobem snímána data o pohybu puku i hráčů, celkově je tak možné sesbírat až 1 milion trojrozměrných souřadnic a datových záznamů během jednoho zápasu [8]. Sesbíraná data dávají pověřeným osobám – trenérům, sportovním ředitelům – hlubší vzhled do zápasového dění a možnost dělat informovaná rozhodnutí ve prospěch svého týmu. V případě zveřejnění, například formou zajímavých vizualizací, mohou být taková data prospěšná i pro všední fanoušky. Analýzu sportovních statistik je možné dělit na popisnou, předvídající a předepisující [9]. Tato práce se bude zabývat zejména první jmenovanou kategorií.

Cílem práce je vytvoření webové prezentace statistik českých hokejistů v NHL. Cílem práce je počáteční rešerše webových stránek, které se zabývají vizualizací hokejových statistik (především těch pokročilých). Dále také analýza dostupných datových zdrojů a zhodnocení jejich vhodnosti pro účel práce. Cílem práce je prezentace vizualizací, které srozumitelně prezentují pokročilejší statistiky a pohledy na hráče ve formě grafů. Vizualizace se budou zabývat především následujícími aspekty: jak se hráčům dařilo v jednotlivých zápasech, jejich výkonům během celé sezóny a jak se vyvíjí role českých hokejistů jako celku dlouhodobě v NHL.

Kapitola 1

Lední hokej

Aby bylo možné pochopit data a následně se zabývat jejich vizualizací v přehledné a účinné formě, je potřeba nejprve porozumět událostem a akcím, které popisují. Proto se první kapitole věnuje přiblížení podstaty ledního hokeje a jeho pravidlům. Zároveň budou popsány některé speciální situace, které mohou během zápasu nastat.

1.1 Historie

Tento týmový sport, ojedinělý tým, že jeho účastníci klouzají po ledové ploše na nabroušené oceli, má své počátky v kanadském Montrealu. Zde byl 3. března 1875 sehrán první zápas na kryté ploše, podle pravidel převážně převzatých z pozemního hokeje. Přestože k této akci došlo před bezmála 150 lety, některé klíčové aspekty zůstaly nezměněné. Konkrétně se jedná o délku kluzišť a využití puku ve hře.

Na počátku 20. století byly kanadská pravidla hokeje přijaty tehdejší Ligue Internationale de Hockey sur Glace v Paříži, předkem současné mezinárodní organizace IIHF (International Ice Hockey Federation – Mezinárodní federace ledního hokeje). Souběžně se vyvíjel i sport jako takový – začala průmyslová výroba hokejek, součástí hráčské výbavy se staly holenní chrániče a brankáři začali nosit chránič hrudi. Zároveň se napříč Kanadou stavěly zimní stadiony, ty ovšem stále spoléhaly na přírodní led.

Vznik zastřešující organizace IIHF, v roce 1908, dal prostor pro vznik prvních mezinárodních turnajů. Posléze se nový sport představil na letních olympijských hrách 1920, avšak dnes je oblíbenou součástí zimních her. Původně se olympijských her mohli účastnit jen amatérští sportovci, ale to se mění koncem 20. století, kdy byla účast povolena i profesionálům. Hráči z prestižní severoamerické NHL (National Hockey League – Národní hokejová liga) se na hrách poprvé představili roku 1998, v japonském Naganu. [10, 11, 12]

1.2 NHL

Nejvýznamnější ligu ledního hokeje na světě je kanadsko-americká NHL a to téměř ve všech aspektech – ať se jedná o fanouškovskou návštěvnost, kvalitu hráčů nebo celosvětovou popularitu týmů. Zároveň se řadí mezi nejvýznamnější sportovní ligy Severní Ameriky, spolu s celosvětově proslulou basketbalovou NBA (National Basketball Association – Národní basketbalová asociace), baseballovou ligou MLB (Major League Baseball) a ligou amerického fotbalu NFL (National Football League – Národní fotbalová liga).

Její počátek se datuje k roku 1917, kdy se rozpustila kanadská NHA (National Hockey Association – Národní hokejová asociace). To vedlo ke vzniku NHL se čtyřmi týmy. V sezóně 1925/1926

došlo k rozšíření do Spojených států, kdy se soutěže účastnilo až deset týmů. Následně se počet týmů opět snížil a v sezóně 1942/1943 se jednalo už jen o šestici, dnes obecně známou jako „Original Six“. Na další změnu ve složení týmů si fanoušci počkali až do roku 1967, kdy se počet účinkujících týmů zdvojnásobil.

První konkurenci pro hokejovou ligu přinesl rok 1972, kdy vznikla WHA (World Hockey Association – Světová hokejová asociace) rovněž s dvanácti týmy. Výsledkem soupeření těchto dvou organizací byl drastický nárůst v platech hokejistů a také nová liga s názvem NHL, která vznikla spojením obou lig. Původně bylo součástí nové soutěže 21 týmů, ale spolu se vzrůstajícím zájmem fanoušků, přibývalo i stále více týmů. K poslední změně došlo v roce 2021 vznikem týmu Seattle Kraken, již 32. v pořadí. [10, 13]

1.3 Pravidla hry v NHL

Hokejový zápas se skládá ze třetin, kde každá trvá 20 minut čistého času. Jedná se o týmový sport, kde proti sobě v jeden čas nastupují dva šestičlenné týmy, přičemž jedním členem může být na každé straně brankář. Hokej je specifický tím, že se hráči střídají na hrací ploše i během hry. Na nejvyšší úrovni stráví na ledě každý hráč kolem jedné minuty v kuse, což je dáno intenzitou hry. Profesionální týmy mívají celkem 20 hráčů, aktivní bruslaři (mezi ně nepočítáme brankáře) se střídají po čtyřech útočných trojicích a třech obranných dvojicích. Ve speciálních případech se tým může rozhodnout odvolat brankáře a nahradit ho bruslařem, to se typicky stává ke konci zápasu, když tým prohrává a je potřeba vstřelit gól. Jinak brankář bývá na ledě celý zápas. Výjimečně se stává, že dojde během zápasu k výměně brankářů. Na konci zápasu musí být vždy jeden tým vítězný – pokud je po základní hrací době vyrovnáno, zápas pokračuje v rámci tzv. nastaveného času. Nastavený čas je ukončený, jakmile jeden z týmů vstřelí gól a vyhraje tak zápas. Nastavený čas během základní části sezóny trvá pět minut, během kterých má každý z týmů typicky na ledě čtyři hráče – brankáře a tři bruslaře.

Hlavním cílem obou týmů je vstřelit puk do branky soupeře a skórovat tak gól. Hra začíná vhazováním, kdy rozhodčí vhodí puk mezi dva hráče ze soupeřících týmů. Mezi další zajímavé akce na ledě patří střely a případné góly spolu s údajem o asistenci, hity, ztráta a zisk puku nebo vyloučení. K vyloučení dojde, když jeden či více hráčů poruší pravidla hry. Standardní vyloučení trvá dvě minuty hracího času, kdy potrestaný tým hraje v početní nevýhodě. Tento stav trvá dokud nevyprší čas vyloučení, nebo zvýhodněný tým nevstřelí gól. Protože takový trest dává výraznou převahu jednomu z týmů, je nutné ho také zohlednit ve veškeré statistické analýze.

Hrací plocha je rozdělena na tři části dvojicí modrých čar, z pohledu každého z týmů se jedná o obranné, střední a útočné pásmo. Týmy si po konci třetiny vždy vyměňují strany, což je potřeba zohlednit v případě vizualizací dat v prostoru. [14]

Analýza webových zdrojů

Postupem času se rozšiřuje zájem všedních lidí porozumět dění na ledové ploše a zároveň přibývá dat, ke kterým má širší veřejnost přístup. Kombinace těchto faktorů dává prostor pro vznik mnoha zajímavých webů, jejichž autory nejsou lidé spojeni přímo s NHL. Často se jedná o fanoušky sportu se zájmem hlouběji porozumět událostem na ledě a podělit se o tyto informace s podobně smýšlející komunitou. Navzdory tomu stále platí, že nejspolehlivějším zdrojem je přímo web NHL.

2.1 Oficiální stránky NHL

Webové stránky NHL poskytují informace o odehraných zápasech, týmech i jednotlivých hráčích v NHL. V této podkapitole bude popsáno, jaké konkrétní informace je zde možné najít.

Hlavní stránka webu poskytuje statistický přehled pod záložkou *Stats*. Zde je dostupný jednoduchý přehled nejúspěšnějších desíti hráčů v různých kategoriích podle herních pozic – bruslaři,¹ obránci, nováčci a brankáři. Hráči jsou seříděni podle statistik spojených s jejich pozicí, u bruslařů se jedná například o vstřelené branky nebo asistence, u brankářů pak o průměr obdržených gólů za zápas nebo úspěšnost zákroků vyjádřenou v procentech. K dispozici jsou i tabulkové pohledy, kde je možné si dohledat základní statistiky pro veškeré hráče. Ačkoliv jsou informace zde zajímavé, pro potřeby práce nejsou zcela relevantní, protože se jedná převážně právě o tabulky. Pro detailnější statistiky a zejména vizualizace je potřeba navštívit záložku *NHL EDGE Advanced Stats*.

2.1.1 NHL EDGE

Počínaje sezónou 2021/2022 uvedla NHL do provozu technologii *NHL EDGE*, jejíž hlavní funkcí je sbírat prostorová data hráčů a puku během zápasů v každém ze 32 zimních stadionů, kde se hrají zápasy NHL. Tato iniciativa značně rozšířila dostupné informace o zápasech. Kromě možnosti analyzovat, kde se odehrála každá jednotlivá akce na ledě, dostali fanoušci přístup i k doplňujícím datům. Mezi ně patří například rychlost bruslení jednotlivých hráčů, vzdálenost, kterou každý hráč za zápas urazil, nebo rychlost střel. Dnes už jsou takové informace běžně součástí i televizních vysílání. [7]

¹Tento pojem zahrnuje obecně veškeré hráčské pozice, které během zápasu aktivně bruslí, tedy útočníci i obránci. Vyčlenění z této kategorie jsou brankáři.

2.2 Ostatní

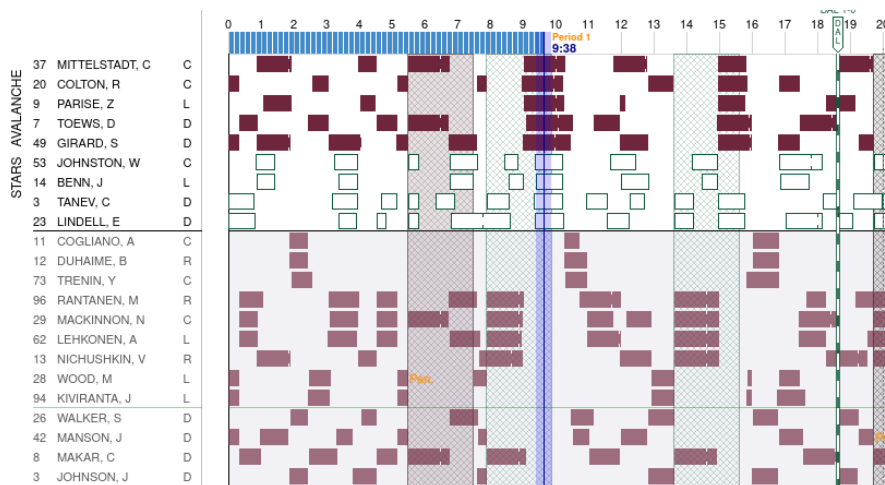
Mimo oficiální stránky existuje mnoho dalších zajímavých zdrojů. Příklady webů v této sekci jsou převážně projekty jednotlivých lidí, kteří statistické přehledy nabízejí volně, nebo případně částečně zpoplatněnou formou.

2.2.1 NHL.CZ

Zdrojem statistik z NHL se zaměřením přímo na české a slovenské hráče je web nhl.cz. Zde se nacházejí jednak týmové soupisky, kde jsou čeští i slovenští hráči zvýrazněni. Zároveň je zde možné najít jejich souhrnné statistiky, jako počet gólů a asistencí, ve formě tabulky. Tabulky se nachází pod záložkou „Češi a Slováci“ v oddělených sekcích „Statistiky Čechů“ a „Statistiky Slováků“. [15]

2.2.2 Shiftchart

Web ShiftChart se zaměřuje na jeden konkrétní aspekt každého zápasu – kdy byl každý z hráčů během zápasu na ledě. Vizualizace se skládá z časové linie daného zápasu, kde je u každého hráče zvýrazněný čas strávený na ledě. Součástí vizualizace jsou i další zápasové situace – speciální označení mají přesilovky a vstřelené góly. Posunem ukazatele na horní liště je zároveň možné zjistit, kteří hráči byli zrovna na ledě v každém momentě zápasu. Dodatečně je možné srovnávat hráče podle celkového času na ledě. Ukázka vizualizace dostupné na tomto webu se nachází na obrázku 2.1. [16]



■ Obrázek 2.1 Ukázka z webu *shiftchart.com* [16]

2.2.3 Hockeyviz

Zdrojem nejunikátnějších vizualizací je web Hockeyviz. Zde je možné najít zejména vizualizace na úrovni týmů. Jednak ve formě souhrnů jednotlivých týmů, ale také jako srovnání dvou týmů proti sobě, což se vztahuje zároveň na náhledy budoucích zápasů a souhrny odehraných zápasů. Vizualizace spojené s jednotlivými hráči pak zobrazují lehce pokročilé statistiky spojené s očekávaným počtem vstřelených gólů. [17]

2.2.4 Interaktivní mapa

Web World Population Review nabízí zejména populační informace pro všechny státy na světě. Kromě toho na této stránce nabízí i informace o národnosti hokejových hráčů v NHL. Data jsou podána formou interaktivní mapy světa, kde přejetím kurzoru po zvoleném státě je možné zjistit počet hráčů původem z daného státu. Ve spodní části stránky jsou detailnější data podána také ve formě tabulky. [18]

Analýza datových zdrojů

Stejně jako v předcházející kapitole, i zde platí, že nejobsáhlejší a nejspolehlivějším zdrojem dat je přímo web NHL. Tato kapitola se zabývá popisem dat, která jsou přístupná veřejnosti, a jejich strukturou.

Oficiální data o jednotlivých zápasech, týmech i hráčích jsou zpřístupněna přes API (Application Programming Interface). Zatím neexistuje žádná oficiální dokumentace tohoto API, ale vznikla spolehlivá neoficiální dokumentace, přístupná v GitHub repozitáři. Veškerá data jsou dostupná ve formátu JSON (JavaScript Object Notation).

3.1 Pořadí týmů

Informaci o pořadí týmů v lize poskytuje API endpoint *standings*.¹ Dotaz na adresu vrací seznam týmů podle pořadí v lize k určitému datu, podle údaje {date} v URL (Uniform Resource Locator) adrese.² Alternativně lze tento parametr nahradit hodnotou *now*, čímž je proveden dotaz na aktuální pořadí. Tímto způsobem lze získat seznam všech aktivních týmů, který slouží k vytváření dalších API požadavků. Například údaj o zkratce týmu lze využít v požadavku pro soupisky daného týmu, což je popsáno v podkapitole 3.4.

```
{..., "standings": [  
  0: {"teamName": { "default": "Florida Panthers", ...}  
    "teamAbbrev": { "default": "FLA", ...}  
    "teamLogo": "https://assets.nhle.com/logos/nhl/svg/FLA_light.svg"  
  }, ...]}
```

■ **Výpis kódu 1** Ukázka záznamu o týmu

3.2 Seznam sezón

Předchozí dotaz umožňuje získat informaci o pořadí týmů v libovolný den některé sezóny. K vytvoření dotazu je však potřeba vědět, kdy sezóny probíhaly. Ke stažení popisných dat o každé

¹ <https://api-web.nhle.com/v1/standings/{date}>

² Datum je vždy ve formátu YYYY-MM-DD

dostupné sezóně lze využít endpoint *standings season*.³ Popisná data obsahují například informace o identifikačním čísle sezóny⁴, datu začátku i konce základní části, ale také o způsobu organizace – zda byly v dané sezóně týmy rozděleny do konferencí, případně do divizí, zda mohl zápas skončit remízou nebo jestli získával poražený tým v nastaveném čase body do celkového pořadí.

3.3 Plán zápasů

Podstatnou informací je plán zápasů na každý den. Podle zápasů a údaje o zúčastněných týmech lze získat soupisky obou týmů a dále pak i jednotlivé hráče. Získání seznamu zúčastněných hráčů je klíčové pro analýzu jejich statistických měřítek z vybraného dne a posléze k vizualizaci jejich výkonů. K tomu slouží endpoint *schedule*,⁵ kde údaj {date} v URL adrese je požadované datum. Dotaz ovšem nevrací data jen pro jeden konkrétní den, nýbrž celý týden, jehož je zvolený den součástí. K získání dat odpovídajících zvolenému dni, je potřeba identifikovat požadovaný záznam ze seznamu podle vyhledávaného data.

```
{
  "nextStartDate": "2024-02-13",
  "previousStartDate": "2024-01-30",
  "gameWeek": [
    0: {
      "date": "2024-02-06",
      "dayAbbrev": "TUE",
      "numberOfGames": 8,
      "games": [
        0: {
          "id": 2023020783,
          "season": 20232024,
          "gameType": 2,
          "awayTeam": {
            "id": 20,
            "placeName": {"default": "Calgary"},
            "abbrev": "CGY",
            "score": 4, ...},
          "homeTeam": {
            "id": 6,
            "placeName": {"default": "Boston"},
            "abbrev": "BOS",
            "score": 1, ...},
          ...},
        ...], ...},
    ...], ...}
  ...], ...}
```

■ Výpis kódu 2 Ukázka záznamu plánu zápasů

³<https://api-web.nhle.com/v1/standings-season>

⁴Běžně základní část NHL trvá od října do dubna následujícího roku, sezóna je tedy identifikována přelomem let, ve kterých probíhala, ve formátu bez rozdělovače, např. „20232024“

⁵<https://api-web.nhle.com/v1/schedule/{date}>

3.4 Soupisky

Jelikož se aplikace soustředí na statistiky jednotlivých hráčů, je třeba získat informace o veškerých hráčích, kteří nastupují za daný tým. K tomuto účelu slouží endpoint *roster*,⁶ kde údaj {team} v URL adrese je nahrazený právě zkratkou vybraného týmu. Výsledná data jsou strukturována jako tři oddělené seznamy hráčů, odlišené podle herní pozice – útočníci, obránci a brankáři.

3.5 Hráči

Detailní informace o jednotlivých hráčích jsou dostupné přes endpoint *player*,⁷ kde údaj {player_id} v URL adrese je identifikační číslo hráče. Dotaz na tento endpoint je prováděný v případě, že je hráčská data nutné filtrovat podle národnosti hráče. Požadavkem je však možné získat i další informace – od výšky a váhy hráče, přes číslo dresu a datum narození, po souhrnné statistiky v posledních pěti zápasech nebo za celou kariéru.

3.6 Střídání

Pro efektivní analýzu výkonu hráče a jeho vlivu na dění v zápase je nutné zjistit, kdy se během zápasu hráč nacházel na ledě. To nebývá zcela jasné, jelikož hokej je dynamická hra, kde dochází ke střídáním často i během hry. Proces je značně zjednodušený tím, že informace o střídáních jsou dostupné přes API endpoint *shiftchart*,⁸ kde je nahrazením parametru {game_id} za identifikační číslo zápasu v URL adrese možné získat záznamy o veškerých střídáních během daného zápasu. Podstatné informace ze získaných dat jsou primárně „startTime“ (čas začátku střídání), „endTime“ (čas konce střídání) a „duration“ (doba trvání).

```
{"data": [ 0: {"id": 13839944,
  "playerId": 8473419,
  "firstName": "Brad",
  "lastName": "Marchand",
  "startTime": "00:00",
  "endTime": "00:44",
  "duration": "00:44",
  "shiftNumber": 1,
  "period": 1,
  "gameId": 2023020783,
  "teamId": 6,
  "teamAbbrev": "BOS",
  "teamName": "Boston Bruins",
  "hexValue": "#111111",
  ...],
"total": 808}
```

■ Výpis kódu 3 Ukázka záznamu o střídáních

⁶ <https://api-web.nhle.com/v1/roster/{team}/current>

⁷ https://api-web.nhle.com/v1/player/{player_id}/landing

⁸ https://api.nhle.com/stats/rest/en/shiftcharts?cayenneExp=gameId={game_id}

3.7 Play-by-play

Základem statistické analýzy na úrovni jednotlivých zápasů jsou tzv. „play-by-play“ data. Jedná se o záznamy o každé významné akci během zápasu, jako je střela, hit nebo vhazování. Kromě informace o zápasovém čase, kdy k akci došlo, a zúčastněném hráči (případně hráčích), lze získat i prostorová data. Souřadnicové údaje s informací o umístění akce na ledové ploše budou později využity v grafu, který vystihuje aktivitu hráče v zápase.

Data jsou dostupná na endpointu *play-by-play*.⁹ Místo údaje {game_id} v URL adrese je doplněno identifikační číslo zápasu.

```
{
  "id": 2023020783, ...,
  "plays": [...,
    1: {
      "eventId": 51,
      "periodDescriptor": {...},
      "timeInPeriod": "00:00",
      "timeRemaining": "20:00",
      "situationCode": "1551",
      "homeTeamDefendingSide": "left",
      "typeCode": 502,
      "typeDescKey": "faceoff",
      "sortOrder": 11,
      "details": {
        "eventOwnerTeamId": 20,
        "losingPlayerId": 8475745,
        "winningPlayerId": 8474150,
        "xCoord": 0,
        "yCoord": 0,
        "zoneCode": "N"}},
    2: {
      "eventId": 103,
      "periodDescriptor": {...},
      "timeInPeriod": "00:09",
      "timeRemaining": "19:51",
      "situationCode": "1551",
      "homeTeamDefendingSide": "left",
      "typeCode": 507,
      "typeDescKey": "missed-shot",
      "sortOrder": 12,
      "details": {
        "xCoord": -16,
        "yCoord": -40,
        "zoneCode": "N",
        "reason": "wide-right",
        "shotType": "wrist",
        "shootingPlayerId": 8474150,
        "goalieInNetId": 8480280,
        "eventOwnerTeamId": 20}}, ...]}
}
```

■ **Výpis kódu 4** Ukázka záznamu o zápasových akcích

⁹https://api-web.nhle.com/v1/gamecenter/{game_id}/play-by-play

3.8 Zápasové záznamy

K analýze statistik jednoho hráče v rámci sezóny slouží API endpoint *game log*¹⁰. Požadavek vrací data ve formě seznamu veškerých zápasů, které tento hráč odehrál během sezóny. Jaké zápasy jsou v seznamu zahrnuty, je specifikováno následujícími parametry v URL adrese – {player_id} reprezentuje identifikační číslo hráče, {season} udává o jakou sezónu se jedná a {game_type} určuje typ zápasů (jedná se o zápasy z přípravy, základní části nebo playoff). Endpoint slouží jako zdroj dat pro vytvoření přehledu statistik jednotlivých hráčů během sezóny, jak je dále popsáno v podkapitole 5.1.4.

```
{"seasonId": "20232024",
"gameTypeId": 2,
"playerStatsSeasons": [...],
"gameLog": [
  0: {
    "gameId": 2023021309,
    "teamAbbrev": "CGY",
    "homeRoadFlag": "H",
    "gameDate": "2024-04-18",
    "goals": 0,
    "assists": 0,
    "commonName": {"default": "Flames"},
    "opponentCommonName": {"default": "Sharks"},
    "points": 0,
    "plusMinus": 1,
    "powerPlayGoals": 0,
    "powerPlayPoints": 0,
    "gameWinningGoals": 0,
    "otGoals": 0,
    "shots": 2,
    "shifts": 20,
    "shorthandedGoals": 0,
    "shorthandedPoints": 0,
    "opponentAbbrev": "SJS",
    "pim": 0,
    "toi": "15:32"},
  ...]}

```

■ **Výpis kódu 5** Ukázka zápasových záznamů

¹⁰https://api-web.nhle.com/v1/player/{player_id}/game-log/{season}/{game_type}

Použité nástroje

Následující kapitola se zabývá konkrétními nástroji a technologiemi, které byly v práci využity. Podstatou kapitoly je popsat jednak hardware, pomocí kterého byl vývoj provedený, a dále programy, programovací jazyky a jiné technologie, využití k realizaci samotné aplikace.

4.1 Hardware

Stahování a úprava dat, stejně jako psaní zdrojového kódu aplikace spolu se skripty na stažení dat bylo prováděno na osobním počítači (notebooku) s následujícími parametry.

```
Výrobce a typ notebooku: Lenovo ThinkPad L490
Procesor: Intel® Core i7-8565U CPU @ 1.80GHz × 8
RAM: 16 GB
Pevný disk: 512 GB
Grafická karta: Mesa Intel® UHD Graphics 620 (WHL GT2)
Operační systém: Pop!_OS 22.04
```

4.2 Software

Veškeré skripty a grafy, jakožto i aplikace samotná, jsou napsané, vytvořené a vyvinuté v jazyce R. R je programovací jazyk a prostředí (run-time environment), které nabízí širokou škálu nástrojů pro statistické výpočty a grafické funkce. Jedná se o svobodný software (free software), jehož zdrojový kód je zveřejněný pod obecnou veřejnou licencí GNU¹ (GNU General Public License). [20]

Programovací jazyk R ve své základní formě poskytuje funkce podporující provedení veškerých zásadních statistických výpočtů, vykreslení grafů a operace s maticemi, mimo jiné. Tyto funkcionality je možné dále rozšířit pomocí specializovaných balíčků. Tímto způsobem se možnosti jazyka R posouvají mnohem dále než je původní statistické zaměření. Mezi nejzajímavější takové balíčky patří *shiny*, jenž rozšiřuje možnosti co se týče webové prezentace dat, nebo *rmarkdown*, který zase slouží k vytváření dynamických dokumentů založených na reprodukovatelných procesech (ve výsledku se podobá Jupyter notebooku pro Python). [21] V práci byly využity následující balíčky:

¹Licence je spojená s projektem GNU (GNU's Not Unix). Software zveřejněný pod touto licencí dává všem jeho uživatelům možnost sdílet a měnit jeho zdrojový kód. Podmínkou je, že výsledný program bude pod toutéž licencí zaručovat identická práva na přístup ke zdrojovému kódu dalším uživatelům. [19]

- *shiny*² – nabízí rámec pro vývoj interaktivních webových aplikací s automatickými reaktivními vazbami mezi vstupem a výstupem,
- *shinyWidgets*³ – nadstavba balíčku *shiny*, poskytuje kolekci komponent uživatelského rozhraní,
- *ggplot2*⁴ – systém pro vykreslování vizualizací a grafů,
- *plotly*⁵ – slouží k vytvoření interaktivních grafů zobrazovaných na webu na základě *ggplot2*, využívá JavaScriptovou knihovnu *plotly.js*,
- *sportyR*⁶ – nadstavba *ggplot2* k vykreslení škálovaných hracích ploch pro různé sporty,
- *plyr*⁷ – sada nástrojů pro časté manipulace s tabulkami – rozdělení datové struktury do podsekcí, aplikování funkce na jednotlivé části a následné spojení zpět do jednoho datového celku,
- *dplyr*⁸ – balíček pro nejčastější manipulace s daty (vytváření nových sloupečků na základě existujících, výběr sloupců podle jmen nebo filtr tabulkových záznamů),
- *httr*⁹ – poskytuje funkce pro vytváření HTTP (Hypertext Transfer Protocol) dotazů, typicky se jedná o GET() a POST(),
- *tidyr*¹⁰ – balíček se zaměřuje na úpravu tabulkových záznamů, týká se to změny dimenze tabulek, spojování/rozdělování sloupců a zacházení s nulovými hodnotami,
- *lubridate*¹¹ – řeší veškeré záležitosti ohledně manipulace a čtení časových záznamů v R,
- *mongolite*¹² – MongoDB klient pro R.

Zdrojem veškeré dokumentace a zdrojových kódů týkajících se jazyka R, stejně jako jeho balíčků, jsou webové stránky CRAN¹³ (The Comprehensive R Archive Network – Obsáhlá síť archivů R). Jedná se o „*síť FTP¹⁴ a webových serverů po celém světě, schraňující identické a aktuální verze zdrojového kódu a dokumentace pro jazyk R*“ [22].

4.2.1 RStudio

K vytvoření skriptů pro stažení dat i vývoji samotné aplikace byl využitý program RStudio, IDE (Integrated Development Environment – vývojové prostředí) pro programování v jazyce R. Program existuje ve dvou verzích – RStudio Desktop je desktopová aplikace, ve variantě RStudio Server pak software běží na vzdáleném serveru a RStudio je přístupné přes webový prohlížeč. Rstudio také podporuje programování v jazyce Python. [23]

²Dokumentace *shiny* zde: <https://cran.r-project.org/web/packages/shiny/index.html>

³Dokumentace *shinyWidgets* zde: <https://cran.r-project.org/web/packages/shinyWidgets/index.html>

⁴Dokumentace *ggplot2* zde: <https://cran.r-project.org/web/packages/ggplot2/index.html>

⁵Dokumentace *plotly* zde: <https://cran.r-project.org/web/packages/plotly/index.html>

⁶Dokumentace *sportyR* zde: <https://cloud.r-project.org/web/packages/sportyR/index.html>

⁷Dokumentace *plyr* zde: <https://cran.r-project.org/web/packages/plyr/index.html>

⁸Dokumentace *dplyr* zde: <https://cloud.r-project.org/web/packages/dplyr/index.html>

⁹Dokumentace *httr* zde: <https://cran.r-project.org/web/packages/httr/index.html>

¹⁰Dokumentace *tidyr* zde: <https://cloud.r-project.org/web/packages/tidyr/index.html>

¹¹Dokumentace *lubridate* zde: <https://cloud.r-project.org/web/packages/lubridate/index.html>

¹²Dokumentace *mongolite* zde: <https://cran.r-project.org/web/packages/mongolite/index.html>

¹³Dostupné zde: <https://cran.r-project.org/>

¹⁴File Transfer Protocol

4.2.2 MongoDB

MongoDB se řadí do skupiny databázových systémů s názvem NoSQL („non SQL“ nebo „not only SQL“ – „nejen SQL“). Jedná se o nerelační databáze, které se od těch relačních liší způsobem ukládání záznamů. Relační databáze ukládají data v tabulkovém formátu podle datového modelu a obsahují vazby mezi jednotlivými tabulkami. Oproti tomu nerelační databáze používají systém, který je více flexibilní – mezi typické metody patří ukládání dat ve formátu dvojic klíč-hodnota, ve formě JSON dokumentů nebo jako grafy s vrcholy a hranami. MongoDB používá vlastní verzi schématu pro ukládání dat odvozenou od systému JSON dokumentů, nazvanou BSON (Binary JSON – binární JSON). Typický dokument obsahuje veškerá data týkající se jedné doménové entity bez vazeb na jiné dokumenty. Dokumentová uložiska navíc nevyžadují, aby každý záznam měl stejnou strukturu, což umožňuje větší flexibilitu ve zpracování různorodých dat. [24, 25]

4.2.3 Docker

Docker je platforma v podobě svobodného softwaru (open source), jejímž cílem je zjednodušení procesu vývoje, nasazení a spouštění aplikací. Za tímto účelem využívá tzv. „kontejnery“ – jedná se o „jednotku softwaru, která zahrnuje zdrojový kód, spolu s ostatními nutnými zdroji pro spuštění (systémové nástroje, knihovny a nastavení), aby aplikace běžela rychle a spolehlivě v různých výpočetních prostředích.“ [26] Důsledkem je jednotné chování aplikace bez ohledu na prostředí, ve kterém je spuštěna. [27]

Kapitola 5

Zpracování dat

Následující kapitola se věnuje procesu přípravy dat do formy vhodné pro vytvoření vizualizací. V rámci kapitoly bude popsán nejprve postup stažení dat, včetně jejich úpravy. Druhá část kapitoly přiblíží metody čištění a následného sjednocení dat do tabulek tak, aby mohla být uchována v databázi.

5.1 Stahování dat

Veškeré stahování dat bylo provedeno v rámci skriptů v jazyce R. Jako primární zdroj dat sloužilo API vystavené přímo NHL, blíže popsané v kapitole 3. Na API byly prováděny dotazy pomocí funkce *GET* z balíčku *httr*.¹

5.1.1 Čeští hráči

Základem skriptu pro stažení dat o českých hokejistech je dotaz na NHL API endpoint *rosters*, blíže popsáný v podkapitole 3.4. Jedním z parametrů požadavku je třípísmenná zkratka týmů, jejichž seznam je nejprve potřeba extrahovat z požadavku na endpoint *standings*, kterému se věnuje podkapitola 3.1. Tímto způsobem jsou získána data ve formě tří rozdílných seznamů pro každý tým, kde jsou hráči seskupeni podle pozice – útočníci, obránci a brankáři. Záznamy neobsahují údaj o týmu, za který hráč aktuálně hraje, což je podstatná informace pro vytvoření srozumitelných grafů. K jednotlivým záznamům jsou přidány kolonky s údaji o názvu týmu hráče, získané v předchozím kroku.

Dále je potřeba data uspořádat do jednotné formy, jelikož záznamy pro jednotlivé hráče se často zásadně liší v pojmenování sloupečků. V prvním kroku byla využita funkce *rbind.fill*.² Výsledkem je značně rozšířená tabulka, co se týče počtu sloupců. Uložená data však nejsou sjednocená – například údaje o jméně, místě narození a herní pozici se nevyskytují vždy ve stejných sloupcích, jak je vidět na obrázku 5.1.

Posledním krokem je úprava dat, která byla provedena převážně funkcemi *mutate*³ pro úpravu

¹Funkce přijímá jako vstupní parametr URL adresu, na které vyhodnocuje dotaz typu GET v rámci HTTP (Hypertext Transfer Protocol) protokolu.

²Smyslem funkce *rbind.fill* je sjednotit datové tabulky s některými odlišnými sloupečky. Výsledná tabulka obsahuje sjednocení sloupečků obou vstupních tabulek. Údaje ve sloupcích, které nebyly součástí původní tabulky, jsou doplněny nulovou hodnotou. Funkce je součástí balíčku *plyr*.

³Funkce *mutate* slouží k vytvoření/úpravě/odstranění sloupečků datové tabulky. Typicky se používá pro vytvoření nových sloupečků, které se odvíjí od jednoho nebo více existujících sloupců. Funkce je součástí balíčku *plyr*.

id	default	default.1	firstName.default	lastName.default	default.2	birthCity.default	birthCountry
8482093	Seth	Jarvis	NA	NA	Winnipeg	NA	CAN
8480829	Jesper	Kotkaniemi	NA	NA	NA	Pori	FIN
8475744	NA	NA	Evgeny	Kuznetsov	NA	Chelyabinsk	RUS
8477962	Brendan	Lemieux	NA	NA	Denver	NA	USA
8476921	Jordan	Martinook	NA	NA	Brandon	NA	CAN
8484203	Bradly	Nadeau	NA	NA	St-Francois de Madaw	NA	CAN
8480039	Martin	Nove Mesto na Morave	NA	Necas	NA	NA	CZE
8476474	Stefan	Noesen	NA	NA	Plano	NA	USA
8473533	Jordan	Staal	NA	NA	Thunder Bay	NA	CAN
8480830	Barnaul	NA	Andrei	Svechnikov	NA	NA	RUS
8476882	Teuvo	NA	NA	Teravainen	NA	Helsinki	FIN

■ **Obrázek 5.1** Ukázka formátu uložených informací o hráčích

sloupečků a *unite*⁴ pro sjednocení křestních jmen a příjmení do sloupečku s celým jménem. Na závěr jsou data filtrována, aby ve výsledné tabulce byli pouze hráči narozeni v České republice. To je provedeno podle údajů ve sloupci „birthCountry“ (stát narození).

5.1.2 Zápasové akce

Aby bylo možné vykreslit místa, kde se střely a různé další zápasové akce odehrály na ledové ploše, bylo potřeba stáhnout záznamy o každé akci během zápasu, tzv. „play-by-play“ data. Samotné stažení bylo provedeno ve dvou krocích. Nejprve bylo potřeba získat seznam zápasů z dnešního hracího dne. Za tímto účelem byl proveden požadavek na NHL API endpoint *schedule*, popsáném v podkapitole 3.3. Dotaz vrací seznam záznamů pro každý den v týdnu, který zahrnuje požadované datum. Pro další práci stačí získat jen seznam identifikačních údajů zápasů z vybraného dne.

Následně je potřeba pro každý identifikační údaj zápasu provést požadavek na endpoint *play-by-play*, jemuž se věnuje podkapitola 3.7, který vrací seznam zápasových akcí. Ve výsledné tabulce jsou do řádků sjednocené údaje *gameId* (identifikační číslo zápasu) a *pbp* (seznam náležitostí popisujících akci). Seznam *pbp* je třeba rozbalit do sloupců tabulky, k čemuž slouží funkce *unnest_wider*.⁵ Každá akce tak má v tabulce údaj o čase v zápase, kdy k ní došlo, popis akce a další informace. Ovšem nadále obsahuje ve formě seznamu kolonku *details*, která poskytuje další podstatné informace – zejména souřadnice *x* a *y*, potřebné pro vykreslení akcí na schématu ledové plochy, a identifikátory hráčů. K extrahování hodnot ze seznamu je znovu použita funkce *unnest_wider*. Každý typ akce je specifický tím, že má jinak pojmenované sloupce s identifikátory účastníků se hráčů. Rozdíly v pojmenování jsou zaznamenány v tabulce 5.1.⁶

Z tabulky vyplývá, že záznamy o jednotlivých akcích obsahují identifikátory náležící více hráčům, kteří se akce účastnili. Pro vykreslení akcí, na kterých se každý hráč podílel, je vhodné data modifikovat tak, aby jeden řádek odpovídal jednomu hráči a ne zápasové akci. Za účelem takového rozdělení záznamů byla použita funkce *pivot_longer*.⁷ Některé vzniklé řádkové záznamy jsou ve výsledku nepodstatné a mohou být odstraněny. Ze výsledné tabulky jsou odfiltrovány údaje o brankářích, kteří se účastnili akce typu „střela mimo“ a záznamy o gólových asistencích, nebo-li údaje, kde typ hráče odpovídá *assist1PlayerId* nebo *assist2PlayerId*.

⁴Funkce *unite* se zaměřuje na spojování dat z více sloupečků do jednoho řetězce. Funkce je součástí balíčku *tidyr*.

⁵Funkce *unnest_wider* uchovává řádky, ale modifikuje sloupce. Přirozeně se hodí pro seznamy s pojmenovanými složkami, pak jsou jména uchována v popisku nového sloupce. Funkce je součástí balíčku *tidyr*.

⁶V přehledu typu akcí chybí akce *period-start* (začátek třetiny), *period-end* (konec třetiny), *stoppage* (přerušení hry) a *delayed-penalty* (zpožděný trest), jelikož typicky nemají údaj o zúčastněném hráči ani souřadnicích a nejsou tak pro vizualizaci podstatné.

⁷Funkce *pivot_longer* zvyšuje počet řádků a naopak snižuje počet sloupců datové tabulky. Vstupem je seznam sloupců, které jsou ve výsledku „odstraněny.“ Informační hodnota je však zachována – parametr *names_to* určí název sloupce, kde budou uchovány názvy zrušených sloupců, a identicky *values_to* značí sloupec pro umístění hodnot, které náležely původním sloupečkům. Funkce je součástí balíčku *tidyr*.

■ **Tabulka 5.1** Označení identifikátorů hráčů v „play-by-play“ datech

Kód akce	Typ akce	Hráč 1	Hráč 2	Hráč 3
faceoff	vhazování	winningPlayerId	losingPlayerId	
shot-on-goal	střela na bránu	shootingPlayerId	goalieInNetId	
blocked-shot	zblokovaná střela	shootingPlayerId	blockingPlayerId	goalieInNetId
missed-shot	střela mimo	shootingPlayerId	goalieInNetId	
goal	gól	scoringPlayerId	assist1PlayerId	assist2PlayerId
hit	hit	hittingPlayerId	hitteePlayerId	
penalty	vyloučení	committedByPlayerId	drawnByPlayerId	servedByPlayerId
giveaway	ztráta puku	playerId		
takeaway	získání puku	playerId		

Poslední úpravou je modifikace záznamů popisujících střely. V datech poskytnutých skrze API jsou události rozdělené podle výsledku střely – zblokovaná střela, střela mimo bránu, střela na bránu a gól. Za účelem sjednocení těchto záznamů, dostávají všechny tyto akce označení *shot* (střela) a do nové kolonky s názvem „shotResult“ (výsledek střely) je uložen výsledek (*blocked* (zblokovaná), *missed* (mimo), *on target* (na bránu) nebo *goal* (gól)). Takovým způsobem je možné v tabulce snadno najít veškeré akce typu střela a zároveň je stále odlišit podle výsledku.

5.1.3 Strídání

Získání dat o jednotlivých střídáních během zápasu pro každého z hráčů je poměrně přímočaré, jelikož NHL za tímto účelem vystavila přímo API endpoint *shiftchart*, jehož popisem se zabývá podkapitola 3.6. K provedení dotazu na tento endpoint je potřeba jediný parametr – identifikační číslo zápasu. Endpoint *schedule*, kterému se věnuje podkapitola 3.3, poskytuje, v kombinaci s datem, seznam zápasů pro každý den v aktuálním hracím týdnu. Ze získaných dat lze následně získat jen údaje z vybraného hracího dne tak, že je v seznamu dní v týdnu nalezen záznam s identickým údajem „date“ (datum), jako má vyhledávaný den. Záznam jednoho dne obsahuje kromě údajů „dayAbbrev“ (zkratka dne) a „numberOfGames“ (počet zápasů) také seznam „games“ (zápasy), kde je každému zápasu přidělený identifikátor.

Pro každý ze zápasů je následně proveden dotaz na endpoint *shiftchart* s využitím identifikačních čísel získaných v předchozím kroku. Výsledná data jsou sloučena do jediné tabulky pomocí funkce *rbind*.⁸

V neposlední řadě je třeba změnit formát časových údajů. Ve stažených datech je informace o čase v zápase, kdy střídání začalo a skončilo, poskytnuta ve dvou údajích – číslo probíhající třetiny a uběhlý čas od začátku aktuální třetiny. Pro další zpracování je však důležité znát čas v rámci celého zápasu. Za tímto účelem je k údají o uběhnutém čase ve třetině připočteno 20 minut za každou uplynulou třetinu zápasu. Pro čtení časových údajů ve správném formátu je využita funkce *parse_date_time*, která slouží k načítání textových řetězců reprezentujících časovou stopu do objektů typu POSIXct.⁹

5.1.4 Krátkodobé statistiky

Sekce krátkodobých statistik se zabývá vývojem několika vybraných statistických ukazatelů pro každého českého hráče během sezóny. Takové zápasové záznamy lze získat dotazem na API

⁸Funkce *rbind* spojuje vektory, datové tabulky a matice po řádcích. Jedná se o jednu ze základních funkcí jazyka R.

⁹Funkce *parse_date_time* umožňuje uživateli specifikovat formát vstupního řetězce a vypořádat se tak s různorodými reprezentacemi časové stopy. Je součástí balíčku *lubridate*.

endpoint *game log* 3.8. Parametry dotazu jsou identifikační čísla hráče a vybrané sezóna. Uložení získaných dat do datasetu dále popisuje podkapitola 5.2.1

Dotaz vrací data ve formě seznamu záznamů z každého odehraného zápasu. Tento seznam je různě dlouhý pro každého hráče, protože vždy odpovídá počtu odehraných zápasů. Pozorované údaje se navíc liší mezi bruslaři a brankáři, takže zpracování probíhá ve dvou krocích podle toho, o jaký typ hráče se jedná. V obou případech jsou ke každému zápasu k dispozici tyto informace:

- „teamAbbrev“ – třípísmenná zkratka názvu týmu hráče,
- „opponentAbbrev“ – třípísmenná zkratka názvu soupeřícího týmu,
- „homeRoad“ – přepínač vyjadřující zda byl hráčův tým v zápase domácí *H*, nebo hostující *R*,
- „gameDate“ – datum zápasu,
- „toi“ – čas na ledě.

Výkon bruslaře popisují následující statistiky:

- „goals“ – počet vstřelených gólů,
- „assists“ – počet asistencí,
- „points“ – počet získaných bodů,
- „shots“ – počet střel,
- „shifts“ – počet střídání.

U brankářů jsou pak evidovány tyto údaje:

- „gamesStarted“ – ukazatel vyjadřuje, zda brankář nastoupil v zápase jako první *1*, nebo byl vystřídaný až v průběhu *0*,
- „shotsAgainst“ – počet střel, kterým brankář čelil,
- „goalsAgainst“ – počet obdržených gólů,
- „shutouts“ – ukazatel vystihuje zápasy, kde brankář neobdržel žádný gól hodnotou *1*, zbytek je pak označený hodnotou *0*,
- „savePctg“ – procentuální úspěšnost zákroků.

5.1.5 Dlouhodobé statistiky

Jedním z cílů aplikace je vystihnout roli českých hráčů jako celku v NHL z dlouhodobého pohledu. Je tedy potřeba sesbírat data napříč několika sezónami, aby bylo možné vyobrazit vývoj vybraných metrik v čase. NHL API poskytuje informace z minulých sezón až do přelomu let 1917/1918, ale rozmezí dat bylo třeba omezit, aby mohla vzniknout přehledná a aktuální vizualizace. Jako rozmezí bylo zvoleno moderní období hokeje, konkrétně se jedná o dobu počínající sezónou 1999/2000.

Výkon skupiny hráčů v sezóně lze hodnotit různými měřítky. Následující statistiky byly vybrány na základě kombinace faktorů, zejména dostupnosti dat ve všech sezónách a informativní hodnoty pro vystihnoutí výkonnostní úrovně a střelecké produktivity skupiny hráčů. Jedná se o následující metriky:

- *Počet hráčů* – celkový počet českých hráčů, kteří se během sezóny objevili na soupisce některého týmu,

- *Čas na ledě* – vystihuje lépe roli hráčů v týmu, protože typicky platí, že lepší hráči tráví na ledě více času,
- *Vstřelené góly* – popisuje střeleckou produktivitu skupiny českých hráčů,
- *Získané body* – vystihuje míru, kterou se čeští hráči podíleli na gólových situacích, oproti vstřeleným gólům obsahuje i počet gólových asistencí.

Popisná data o veškerých zaznamenaných sezónách lze získat skrze API endpoint *standings season*, kterému se věnuje podkapitola 3.2. Získaná data slouží ke stažení aktuálního seznamu týmů v každé sezóně, jelikož se počet týmů (nebo jejich jména) v průběhu času mění. Jedná se o údaje „id“ (identifikační číslo sezóny) a „standingsEnd“ (datum konce základní části). Datum dále slouží pro vytvoření dotazu na endpoint *standings* (podkapitola 3.1), který poskytuje data o pořadí týmů v daný den. Tímto způsobem je zaručeno, že výsledný seznam týmů skutečně odpovídá složení týmů ve vybrané sezóně. Pro další krok stačí jediný údaj ze seznamu – „teamAbbrev“ s třípísmenným kódem názvu týmu.

Jednotlivé kódy týmů jsou postupně použity pro stažení týmových soupisek z celé sezóny. K tomu slouží endpoint *roster* z podkapitoly 3.4, jehož druhým parametrem, kromě zkratky názvu týmu, je identifikátor sezóny. Získaná data mají podobu tří rozdílných seznamů, kde jsou hráči rozdělení podle herní pozice. Jedná se o seznamy „forwards“ (útočníci), „defensemen“ (obránci) a „goalies“ (brankáři). Za účelem vytvoření jednotné datové tabulky všech hráčů na soupiskách byly tyto seznamy nejprve zřetězeny a následně jednotlivé záznamy o hráčích spojeny do datové tabulky pomocí funkce *rbind.fill*.¹⁰ Ve výsledné tabulce jsou kromě údajů o hráčích, jako jsou herní pozice, jméno a tým, také informace o jejich původu – městě a státě narození. Tím je možné z výsledného seznamu filtrovat jen české hráče.

Pro české hráče je následně potřeba zjistit statistické záznamy z jednotlivých zápasů. Takové údaje poskytuje API endpoint *game log* (podkapitola 3.8) s parametry odpovídajícími identifikačním číslům hráčů a sezóny. Pro dotazování jsou využity jen identifikátory náležící bruslařům, hráčská data jsou tedy nejprve filtrována pro odstranění záznamů o brankářích. Výsledkem je pro každého hráče seznam se souhrnem statistik z jednotlivých odehraných zápasů. Způsob zpracování získaných dat a jejich uložení je dále popsáno v podkapitole 5.2.4.

5.2 Tvorba datasetů

Stažená data bylo dále potřeba upravit tak, aby byla ve vhodné podobě pro následnou vizualizaci v podobě grafů. To často obnášelo sjednocování dat z různých zdrojů do jedné tabulky nebo různé sloupcové úpravy – ať už odstranění, přidání nebo modifikaci existujících hodnot.

5.2.1 Dataset českých hráčů

Dataset českých hráčů tvoří základ aplikace. Zde jsou uloženy veškeré zásadní informace o každém z hráčů narozeném na území České republiky, kteří jsou hlavním zájmem velké části výsledných vizualizací. Tento dataset také slouží jako zdroj pro následné stahování dalších dat o jednotlivých hráčích, k čemuž slouží především údaj „playerId“ (identifikační číslo hráče). Proces získání základních informací o hráčích byl popsán v podkapitole 5.1.1.

V datasetu jsou uloženy následující informace o každém hráči:

- „playerId“ – identifikační číslo hráče,
- „fullName“ – jméno a příjmení,
- „birthCity“ – město narození,

¹⁰Funkce je blíže popsána v podkapitole 5.1.1

- „birthCountry“ – stát narození,
- „shootsCatches“ – ukazatel dominantní ruky hráče s hodnotami R (pravá), nebo L (levá),
- „sweaterNumber“ – číslo dresu,
- „position“ – hodnota vypovídá o pozici hráče, může být R (pravé křídlo), L (levé křídlo), C (středový útočník), D (obránce), nebo G (brankář),
- „positionGroup“ – zobecnění údaje „position“ o pozici hráče, všechny pozice útočníků nesou označení F ,¹¹ obráncům a brankářům zůstává původní označení D , respektive G ,
- „birthDate“ – datum narození,
- „heightInCentimeters“ – výška v centimetrech,
- „weightInKilogram“ – váha v kilogramech,
- „currentTeamAbbrev“ – třípísmenná zkratka týmu.

Ty jsou doplněny o data popisující krátkodobé statistiky v rámci jedné sezóny. Protože se statistiky liší pro bruslaře a brankáře, je nejprve potřeba rozdělit hráče na dvě tabulky, podle odpovídající pozice. V obou případech je pak proveden iterativní průchod přes kolonku identifikačních čísel hráčů. Identifikační číslo slouží jako parametr v dotazu na zápasové statistiky, jak je popsáno v podkapitole 5.1.4, spolu s identifikátorem sezóny.

Výsledkem dotazu je seznam statistických souhrnů z jednotlivých zápasů. Pro uložení hodnot do datasetu jsou jednotlivá měřítka rozdělena, čímž vznikne několik dílčích seznamů. Každý z těchto seznamů sleduje vývoj jedné proměnné pro vybraného hráče. Vzniknou tak různé seznamy například pro počet gólů, které hráč v zápasech vstřelil, čas strávený na ledě a také ostatní pozorované metriky. Následně je řádek v datasetu českých hráčů, odpovídající danému hráči, rozšířen o sloupceky s vybranými statistikami. Statistické údaje jsou do tabulky vloženy v podobě seznamů. Následně jsou zvláště spojeny řádky odpovídající bruslařům a řádky odpovídající brankářům do dvou oddělených tabulek. Jelikož mají spojované řádky stejné sloupce, je použita funkce *rbind*.¹²

Na závěr jsou sjednoceny tabulky bruslařů i brankářů dohromady. Zde už je použita funkce *rbind.fill*,¹³ kvůli odlišným sloupcům obou tabulek.

5.2.2 Dataset „play-by-play“

Dataset „play-by-play“ slouží k označení míst, kde se na hrací ploše odehrály klíčové akce během zápasů. V kapitole 5.1.2 je popsáno, jakým způsobem jsou data stažena z API a následně upravena, pročištěna a sjednocena. Výsledkem je souhrnný dataset se všemi akcemi na ledě z vybraného hracího dne. Získaná data je však potřeba upravit tak, aby výsledný dataset obsahoval jen záznamy akcí českých hráčů.

Za tímto účelem je využitý dataset českých hráčů, kterému se věnuje podkapitola 5.2.1. Dataset obsahuje údaje o identifikačních číslech hráčů, které lze využít ke spojení tabulky s „play-by-play“ daty pomocí funkce *inner_join*.¹⁴ Tato funkce představuje metodu spojení dvou tabulek na základě identických klíčů ve zvolených sloupcích. Funkce *inner_join* je oproti dalším podobným funkcím¹⁵ specifická tím, že ve výsledné tabulce jsou ponechány jen ty řádky, u kterých byl nalezen párový klíč ve druhé dílčí tabulce. Během operace tak dochází ke ztrátě záznamů,

¹¹ Jsou zde zahrnuti hráči na pozicích levé křídlo, pravé křídlo a středový útočník.

¹² Funkce je blíže popsána v kapitole 5.1.3

¹³ Funkce je blíže popsána v podkapitole 5.1.1

¹⁴ Funkce je součástí balíčku *dplyr*.

¹⁵ V rámci balíčku *dplyr* existují také funkce *left_join*, *right_join* a *full_join*.

u nichž nebyla párová hodnota nalezena. V datasetu tak přetrvávají jen údaje o akcích českých hráčů, zatímco ostatní jsou odstraněny.

Jednotlivé záznamy akcí obsahují mimo jiné prostorové údaje, které odpovídají souřadnicím na ose x (sloupeček „xCoord“) a ose y („yCoord“). S tím souvisí dvě nutné úpravy. Akce typu *střela* jsou vykresleny v grafu jen na jedné polovině ledové plochy (podkapitola 6.1.1) a je tak potřeba provést zobrazení akcí z druhé, nezobrazené části hrací plochy do té zobrazené. Pro ostatní akce pak platí, že nemají jednotnou orientaci. Kvůli střídání stran týmů během zápasů tedy není jasné, na kterou branku hráč útočil, když byla akce zaznamenána. Orientaci akcí na ledě je tak potřeba opravit, aby byl výsledný graf přehledný. Prezencí dat o umístění jednotlivých akcí se zabývá podkapitola 6.1.2.

Transformace souřadnic pro akce typu *střela* je provedena středovou souměrností podle středu hrací plochy. Výsledný graf obsahuje jen kladné body souřadnice x . Pro transformaci hodnot ve sloupečku „xCoord“ je tak využita absolutní hodnota. Nejdříve je ale provedena úprava hodnot souřadnic na ose y – pokud je hodnota souřadnice x („xCoord“) záporná, musí být obráceno znaménko ve sloupečku „yCoord“.¹⁶ K úpravě souřadnicových hodnot ostatních akcí, kromě typu *střela*, je využita kombinace údajů v existujících sloupcích datasetu. Jedná se o kolonky „currentTeamAbbrev“ (zkratka týmu hráče), „homeTeamAbbrev“ (zkratka domácího týmu) a „homeTeamDefendingSide“¹⁷ (umístění obranné třetiny domácího týmu). Je tak možné odvodit u každé akce, na kterou stranu hráč právě útočil. Následně je provedena úprava veškerých souřadnicových dat do jednotné podoby. Orientace hřiště byla určena tak, aby tým hráče měl svou obrannou třetinu vlevo a útočnou vpravo. Souřadnice akcí, které vybranému schématu neodpovídaly, byly transformovány podobně, jako v předchozím kroku akce *střela* – tedy podle středové souměrnosti ve středu hřiště.

Na závěr je potřeba vytvořit ve výsledné tabulce kolonky „eventLabel“ a „shotResultLabel“ s českým překladem existujících sloupečků s typy akcí („event“) a typy střel („shotResult“). Nové kolonky pak slouží jako popisky v legendě výsledného grafu. Způsob překladu vystihují tabulky 5.2 a 5.3.

■ **Tabulka 5.2** Typy akcí

„event“	„eventLabel“
faceoff	vhazování
hit	hit
shot	střela
blocked-shot	zablokovaná střela
committedPenalty	faul
drawnPenalty	faul proti
giveaway	ztráta puku
takeaway	zisk puku

■ **Tabulka 5.3** Typy střel

„shotResult“	„shotResultLabel“
on target	na bránu
goal	gól
blocked	zablokovaná
missed	mimo bránu

5.2.3 Dataset času na ledě

Dataset času na ledě obsahuje údaje o tom, kolik času strávil každý český hráč se všemi ostatními hráči na ledě, jak se spoluhráči, tak i protihráči. Taková informace nelze získat přímo, ale je potřeba provést výpočet pomocí dvou dříve získaných datasetů. Dataset „play-by-play“ z podkapitoly 5.2.2 obsahuje záznamy o všech zápasových akcích z vybraného dne spolu s údaji o hráčích, kteří se akce účastnili. Z tohoto datasetu lze získat všechny české hráče, kteří nastoupili k zápasu (byla k nim připsána během hracího dne alespoň jedna akce). Ze seznamu je potřeba

¹⁶Účelem operace zachování orientace umístění střely vůči brance. Pokud by byly akce čistě překlomeny podle absolutní hodnoty na ose x , znamenalo by to, že střela, která se původně nacházela nalevo od branky, by se ve výsledku nacházela vpravo.

¹⁷Hodnoty jsou dvojího typu – *left* (vlevo) nebo *right* (vpravo).

odstranit záznamy o brankářích, jelikož jejich čas na ledě bývá nepoměrně vyšší než u běžných hráčů. Zahrnutí brankářů do datasetu by výsledný graf značně zkreslila.

Pro výpočet společného času na ledě s ostatními hráči, je nutné pro každý získaný záznam iterativně provést následující kroky:

1. rozdělení záznamů o střídáních¹⁸ na dvě části – střídání jen zvoleného českého hráče a všech ostatních,
2. iterace přes každého z ostatních hráčů, jejichž čas na ledě spolu s českým hráčem bude dále vypočítaný,
3. iterativní průchod údaji o střídáních vybraného českého hráče,
4. ve vnořeném cyklu iterace přes střídání jednoho z ostatních hráčů,
5. výpočet doby, po kterou se střídání českého a jiného hráče překrývají,
6. přičtení času z předchozího kroku k výslednému času, který představuje celkovou dobu, kdy se na ledě vyskytovali oba hráči.

Výsledkem je kompletní tabulka s přehledem společného času na ledě českých hráčů spolu s ostatními hráči, kteří se účastnili zápasu. Vzniklý dataset poslouží jako podklad pro vizualizaci v podkapitole 6.1.3.

5.2.4 Dataset dlouhodobých statistik

V podkapitole 5.1.5 byl popsán proces pro získání seznamu statistických souhrnů ze zápasů každého hráče. Tyto statistiky z jednotlivých zápasů musí být následně shromážděny do souhrnných statistik. Metriky čas na ledě, vstřelené góly a získané body jsou pro každého hráče sloučeny ve dvou formách – jako průměrná hodnota na zápas a jako suma napříč všemi zápasy. Výsledné hodnoty musí být dále shrnuty, aby bylo možné zhodnotit české hráče jako celek v dané sezóně. Znovu jsou použity metody průměrování a sumárního součtu tak, aby výsledný dataset obsahoval zmíněné metriky ve dvou podobách. Dataset tak obsahuje jednak průměrné hodnoty zvolených metrik na hráče, ale také součet hodnot pro všechny hráče společně v jednotlivých sezónách.

Výsledkem je řádek záznamů, kde každá z vybraných statistik je shrnuta do jednoho čísla. Spojením dat z vybraných sezón vznikne výsledný dataset, kde jeden řádek odpovídá jedné sezóně. V datasetu jsou evidovány tyto informace:

- „seasonId“ – identifikační číslo sezóny,¹⁹
- „endDate“ – datum konce základní části,
- „skaters“ – seznam identifikačních čísel všech českých bruslařů v sezóně,
- „goalies“ – seznam identifikačních čísel všech českých brankářů v sezóně,
- „playerCount“ – počet českých hráčů v sezóně (součet délek seznamů „skaters“ a „goalies“),
- „season“ – rok reprezentující sezónu,²⁰
- „toi_sum“ – celkový čas na ledě všech českých bruslařů,
- „toi_avg“ – průměrný čas na ledě v zápase pro české bruslaře,

¹⁸Stažení těchto dat se věnuje podkapitola 5.1.3

¹⁹Sezóny v NHL probíhají typicky na přelomu dvou let, identifikátor obsahuje tyto roky zřetězené za sebou – např. „20232024“ pro sezónu 2023/2024.

²⁰Hodnota odpovídá vždy prvnímu z roků v údaji „seasonId“.

- „goals_sum“ – součet počtu gólů všech českých bruslařů,
- „goals_avg“ – průměrný počet gólů českého bruslaře,
- „points_sum“ – součet počtu bodů českých bruslařů,
- „points_avg“ – průměrný počet bodů českého bruslaře.

Kapitola 6

Vizualizace

Následující kapitola se zabývá využitím dat, jejichž stažení a zpracování se věnuje kapitola 5. V rámci kapitoly budou přiblíženy technologie a metody, pomocí kterých byly vytvořeny grafy prezentující výkony českých hráčů. Podstatou kapitoly je také popsat typy grafů zvolené v jednotlivých vizualizacích a objasnit, proč byly dané typy zvoleny.

6.1 Dnešní dění

Tato sada grafů se zaměřuje na jednotlivé zápasy. Cílem vizualizací v této sekci je představit výkony českých hráčů v jednom zápase, který se odehrál v dnešní den.¹ Podkladem pro vizualizace v této podkapitole jsou zápasová data odpovídající hráčům, kteří se ve zvolený hrací den účastnili některého ze zápasů. Cílem těchto vizualizací je vytvořit pro uživatele přehled, ze kterého bude zřejmá míra zapojení hráče v daném zápase, jeho střelecká produktivita a v neposlední řadě také shrnutí, se kterými ostatními hráči strávil nejvíce času na ledě. Uživatelským vstupem je celé jméno bruslaře, jehož statistiky budou zobrazeny. Uživatel provádí výběr ze seznamu všech českých bruslařů, kteří se účastnili dnešních zápasů.

6.1.1 Schéma střel

Vizualizace s názvem „Schéma střel“ má za cíl vystihnout střeleckou produktivitu jednotlivých českých hokejistů v zápasech z dnešního dne. Jako datový podklad pro vykreslení jednotlivých střel slouží dataset „play-by-play“, jemuž se blíže věnuje podkapitola 5.2.2. Dataset obsahuje záznamy o veškerých zápasových akcích ze zvoleného dne, spolu s údajem o hráči, který se akce účastnil. Data jsou filtrována podle údaje „event“, značícího o jaký typ akce se jedná. Vybrány jsou jen záznamy s hodnotou *shot*.

Jedná se o prostorová data – v kolonkách „xCoord“ a „yCoord“ jsou uloženy hodnoty odpovídající souřadnicím na osách x a y hrací plochy. Výsledkem je *bodový graf*, kde je umístění bodů určeno právě souřadnicemi na osách. Vizualizací podkladem pro vzniklý graf je grafika představující hokejové hřiště, vykreslená pomocí funkce `geom_hockey`.² Funkce přijímá ve vstupním

¹Definice „dnešního dne“ zde není tak přímočará, jak by se mohlo zdát. Jelikož zápasy se typicky hrají ve večerní době severoamerického času, běžně se stává, že probíhají až následující den středoevropského času. Cílem aplikace je pro fanoušky v Evropě (zejména v České republice) prezentovat další den ráno statistiky ze zápasů, které se odehrály během noci. Z tohoto důvodu jsou v sekci *Dnešní dění* používána data, která technicky odpovídají předchozímu dni.

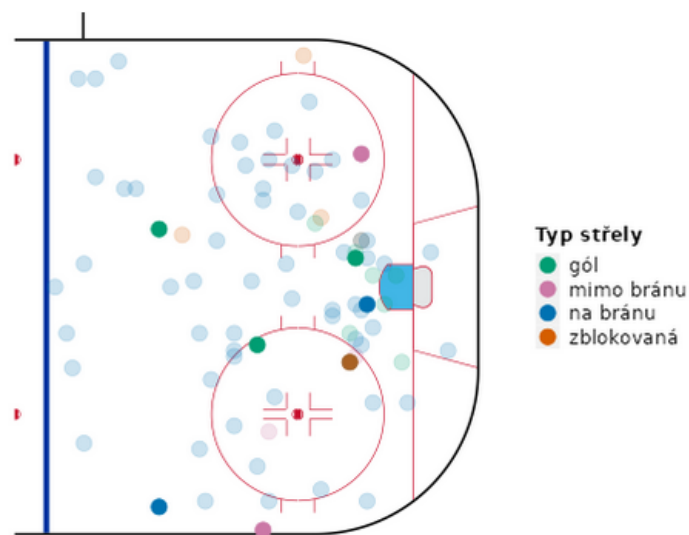
²Funkce `geom_hockey` je součástí balíčku `sportyR`.

parametru název ligy³ a rozmezí hracího pole k vykreslení.⁴ Podle druhého zmíněného parametru lze vykreslit buď celou plochu, nebo jen některou její část. Pro vizualizaci umístění akcí typu střela bylo zvoleno rozmezí parametrem „ozone“ – výsledkem je pouze pravá část hracího pole.

Omezení části ledové plochy k vykreslení bylo zvoleno kvůli lepší přehlednosti dat. Ledová plocha je symetrická – každá akce má odpovídající obraz ve druhé polovině hřiště, podle středové souměrnosti se středem ve středu ledové plochy. Zobrazení všech střel v jedné polovině hřiště přináší tu výhodu, že je jednoduché porovnat umístění střely s ostatními záznamy v datasetu. Oproti vykreslení na celém hřišti, kde je možné střely porovnávat jen s dalšími střelami ve stejné části ledové plochy, což vede na rozpůlení datasetu.

Střely, které jsou v grafu vykresleny, záleží na pozici hráče, zvoleného na uživatelském vstupu. V případě, že se jedná o obránce nebo útočníka, jsou v grafu zobrazeny střely všech českých bruslařů ve vybraný den. Body tak představují záznamy v datasetu „play-by-play“, kde je hráč označený jako „střílející hráč“. Pokud je však zvoleným hráčem brankář, vykreslené body v grafu představují střely, kterým čeští brankáři ve zvolený den čelili. Tyto záznamy pak v datasetu nesou označení „brankář“. V obou případech jsou pak body týkající se přímo zvoleného hráče neprůhledně, zatímco střely ostatních hráčů ve skupině mají zvýšenou průhlednost. Vykreslené body jsou dále barevně odlišeny podle údaje v kolonce „shotResult“, odpovídajícímu výsledku střely. Pro vyjádření významu jednotlivých barev v legendě grafu je využitý sloupeček „shotResultLabel“ s českými popisky, odpovídajícími kolonce „shotResult“. Možné hodnoty popisek jsou *gól*, *na bránu*, *zblokovaná* a *mimo bránu*.

Graf poskytuje přidávaný kontext ohledně střelecké produktivity hráče oproti jen počtu střel nebo gólů. Přestože tyto hodnoty do jisté míry také vypovídají o výkonu hráče v ofenzivní fázi hry, není možné tímto způsobem vystihnout kvalitu jednotlivých střel. Střely v blízkosti soupeřovi branky přinášejí lepší šanci na skórování oproti střelám z velké vzdálenosti. Pomocí grafu tak lze například posoudit schopnost hráče dostávat se do nebezpečných míst pro soupeře a také jeho efektivitu při zužitkování vytvořených šancí.



■ **Obrázek 6.1** Přehled střel ze dne 8. 2. 2024, zvolený hráč Martin Nečas

³Kromě „NHL“ je možné vykreslit hrací plochu s rozměry odpovídajícími různým standardům: „AHL“ (American Hockey League - Americká hokejová liga), „IIHF“ (International Ice Hockey Federation - Mezinárodní federace ledního hokeje), „NWHL“ (National Womens Hockey League - Národní ženská hokejová liga) a další.

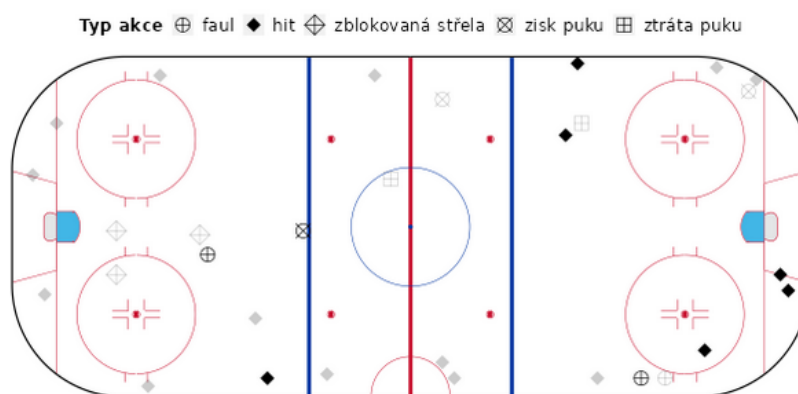
⁴Možnosti vstupu jsou „full“ (celé), „in_bounds_only“ (pouze hrací plocha), „ozone“ (útočná třetina) a další.

6.1.2 Schéma akcí

Veškeré zápasové akce, vyjma střel, jsou vykresleny ve vizualizaci „Schéma akcí“. Datovým podkladem pro graf je v tomto případě, stejně jako pro „Schéma střel“, dataset „play-by-play“, popsáný v podkapitole 5.2.2. V datasetu lze najít záznamy o každé akci, ke které došlo během zápasů daného hracího dne, spolu s údajem o zúčastněném hráči. Z datasetu je nejprve potřeba odstranit záznamy o střelách. Jedná se o takové záznamy, které mají ve sloupečku „event“ hodnotu *shot*. Po provedení filtrace zůstávají tyto typy akcí: *hit* (hit), *faceoff* (vhazování), *takeaway* (zisk puku), *giveaway* (ztráta puku), *blocked-shot* (zablokovaná střela), *committedPenalty* (faul) a *drawnPenalty* (faul proti). Ze zmíněného seznamu je dále potřeba před vykreslením odstranit záznamy o vhazování. Takové akce nejsou zajímavé pro vizualizaci prostorových dat, protože se vždy odehrávají na stejných několika místech na ledové ploše – v kruzích pro vhazování. Jejich zahrnutí v grafu by tak nemělo význam.

Výsledkem je *bodový graf*, kde jsou jednotlivé akce umístěny podle souřadnicových hodnot v kolonkách „xCoord“ a „yCoord“, odpovídajících osám x a y na hrací ploše. K vykreslení grafiky ledové plochy, tvořící podklad výsledného grafu, je použita funkce *geom_hockey*, již dříve zmíněna v podkapitole 6.1.1. Jako vstupní parametr pro určení rozmezí plochy k vykreslení je tentokrát zvolena hodnota „in_bounds_only“, což znamená vykreslení celé ledové plochy. Rozdíl oproti předchozí vizualizaci spočívá v tom, že u akcí kromě střely může jejich přesné umístění na hrací ploše vést k lepšímu vystihnutí situace, během které k akci došlo, spolu s přesnějším vystihnutím role zúčastněného hráče. V datasetu jsou souřadnice akcí zaznamenány tak, aby byly orientovány stejným směrem – tedy obranná třetina je umístěna vlevo a útočí se směrem doprava.

Jednotlivé body v grafu umožňují odvodit kontext akce podle jejich umístění na ledové ploše, podobně jako u schématu týkajícího se střel. Například ke zablokovaným střelám nejčastěji dochází v obranné části hřiště, kde mají i největší význam. Zablokované střely v jiných částech už tak důležité nejsou, jelikož se jedná o méně nebezpečné střely. Dále faul v útočném pásmu je obecně považovaný za zbytečný, oproti faulu v obranném pásmu. Dalším příkladem jsou akce typu hit. Pokud k hitu došlo v obranném pásmu, dá se usuzovat, že soupeřící tým byl v držení puku ve výhodné pozici a hráč se soupeře pokusil zastavit. Oproti tomu hit v útočném pásmu je známkou aktivní týmové obrany („forechecking“) – snahy vyvinout na soupeře tlak při rozehrávce a dostat tak svůj tým do výhodné pozice. Podobně ztráta puku v obranném pásmu je daleko větší chyba, než v útočném pásmu. Jelikož po ztrátě puku v obranném pásmu se typicky dostává soupeřící tým do nebezpečné pozice pro skórování.



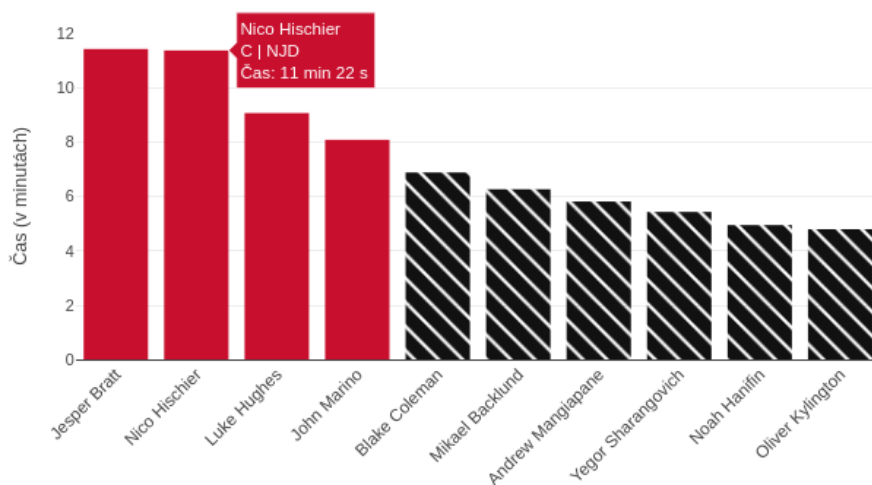
■ **Obrázek 6.2** Přehled akcí ze dne 8. 2. 2024, zvolený hráč Jakub Lauko

6.1.3 Čas na ledě

Graf „Čas na ledě“ má za cíl prezentovat přehled soupeřů i spoluhráčů vybraného hráče a porovnat, kteří z nich strávili na ledě nejvíce času v zápase spolu s českým hráčem. Tímto způsobem lze vyzorovat roli zvoleného hráče v týmu – například když často nastupuje ve formaci s nejlepšími hráči svého týmu, může to svědčit o vysoké úrovni jeho hry. Tento efekt lze pozorovat i z opačného pohledu, tedy když je zkoumán čas na ledě s hráči ze soupeřícího týmu – pokud vybraný hráč nastupuje proti špičkovým hráčům soupeře, svědčí to o jeho spolehlivosti a schopnosti se jim vyrovnat. V grafu chybí společné časy spolu s brankáři, protože ty bývají typicky velmi vysoké oproti ostatním hráčům a výsledný graf by tím byl zkreslován.

Stažení a úpravě dat, které tvoří podklad grafu, se věnuje podkapitola 5.2.3. Jeden řádek v tabulce vyjadřuje, kolik času v zápase spolu strávila dvojice hráčů na ledě – jedná se o českého hráče a jednoho ze všech ostatních hráčů, kteří v zápase nastoupili. V kolonce „tímé“ lze najít společný čas v sekundách. Tato hodnota je pak při vykreslení grafu převedena na minuty. Každý český hráč má v tabulce několik záznamů podle toho, s kolika hráči se vyskytoval na ledě.

Vizualizace dat je provedena *sloupcovým grafem*, kde na ose x jsou vyneseni ostatní hráči a na ose y čas na ledě spolu s českým hráčem, zvoleným uživatelem. Sloupce jsou seřazeny sestupně od hráče s největším společným časem až po hráče s nejmenším. Jedním z uživatelských vstupů je skupina hráčů, kteří jsou v grafu znázorněni. Při zvolení hodnoty *Všichni* jsou v grafu vyneseni všichni hráči bez ohledu na tým. Týmová příslušnost je pak odlišena barvami sloupců. Dalšími možnostmi výběru jsou *Spoluhráči* a *Soupeři*, v každém případě jsou pak údaje ostatních hráčů filtrovány podle kolonky „currentTeamAbbrev“ (zkratka týmu), buď na hodnoty odpovídající týmu zvoleného hráče, nebo soupeřícímu týmu. Druhým uživatelským vstupem je posuvník, jehož hodnota určuje kolik hráčů z vybrané skupiny bude v grafu znázorněno.



■ **Obrázek 6.3** Graf „Čas na ledě“ ze dne 8. 2. 2024, zvolený hráč Ondřej Palát

6.2 Krátkodobé trendy

Podstatou kategorie vizualizací s názvem „Krátkodobé trendy“ je představit uživateli dosavadní výkony českých hráčů v probíhající sezóně. Cílem je prezentovat statistiky jednotlivých českých hráčů v rozsáhlejší měřítku, čímž je možné zasadit statistiky z jednotlivých zápasů, popsanych v podkapitole 6.1, do širšího kontextu.

Jako podklad pro grafy v této sekci slouží dataset českých hráčů, jehož vytvoření se věnuje podkapitola 5.2.1. Záznamy z datasetu obsahují údaje o veškerých českých hráčích, kteří na-

stoupili během probíhající sezóny do zápasu, spolu se záznamy jejich statistik z jednotlivých zápasů ve formě seznamů. V datasetu jsou dále uloženy doplňující informace o každém ze zápasů hráče. Jedná se o údaje „gameDate“ (datum zápasu), „teamAbbrev“ (zkratka názvu týmu), „opponentAbbrev“ (zkratka názvu soupeřícího týmu), „homeRoad“ (ukazatel, zda byl hráčův tým v zápase domácí) a údaj o kolikátý zápas v pořadí se jedná. Tím je možné odvodit další souvislosti ohledně zápasu a výkonu hráče.

V datasetu jsou zaznamenány statistiky bruslařů i brankářů, přestože se jedná o dvě naprosto oddělené skupiny proměnných. Jedinou výjimkou je *Čas na ledě*, který je zaznamenávaný u obou skupin stejným způsobem. Avšak i zde existuje zásadní rozdíl, co se týče měřítka – zatímco brankáři běžně stráví na ledě celý zápas (což odpovídá 60 minutám hracího času), u bruslařů se údaj pohybuje nanejvýš kolem 30 minut. Kvůli zásadním rozdílům v druzích pozorovaných statistik, byla kategorie rozdělena do dvou grafů podle skupiny hráčů. Typově se jedná o stejné grafy, které zobrazují data stejným způsobem. Liší se jen tím, jaký typ hráčů zobrazují a tím pádem i statistikami, které každý z grafů prezentuje.

V obou případech je vykreslována změna hodnoty proměnné v čase, k čemuž se nejvíce hodí *spojnicový* typ grafu. Hodnoty na ose x odpovídají číslu zápasu v pořadí, na ose y se pak jedná o hodnotu pozorované proměnné v daném zápase. Data vykreslená v grafech jsou definovaná dvojicí vstupních parametrů uživatele. Jedná se o dvě výběrové položky – jméno hráče a pozorovaná proměnná. Vstupní parametr se jménem určí hráče, jehož statistiky jsou zvýrazněny. Ve výsledném grafu je vždy vyznačena vybraná statistika pro každého z českých hráčů, ale spojnice odpovídající vybranému hráči je zvýrazněna. Body na zvýrazněné spojnici, odpovídající jednotlivým zápasům, pak obsahují dodatečné informace o zápase. Z popisků těchto bodů lze zjistit datum zápasu, za jaký tým hráč nastoupil a také název soupeřícího týmu. V případě brankářů jsou zvýrazněny některé výjimečné zápasy, v závislosti na pozorované statistice. Mohou to být zápasy, kde byl brankář vystřídán až během hry (nemá v zápase tzv. „start“), nebo kde neoddržel žádný gól. Druhý vstupní parametr určuje, která statistika je v grafu vykreslena. U bruslařů jsou na výběr tyto možnosti:

- *Čas na ledě* – určitým způsobem se jedná o měřítko postavení hráče v týmu. V rozmezí několika zápasů lze sledovat, jak se vyvíjí role hráče.⁵ Příkladem je obrázek 6.4.
- *Góly* – počet gólů hráče v sezóně, hodnota odpovídá kumulativnímu součtu (tedy součet všech gólů od začátku sezóny až po daný zápas).
- *Asistence* – počet asistencí hráče v sezóně, v podobě kumulativního součtu.
- *Body* – počet bodů (góly a asistence) získaných hráčem v sezóně, hodnoty odpovídají kumulativnímu součtu.
- *Střely* – počet střel v zápase. Hodnota poskytuje přehled o aktivitě hráče v zápase a jeho schopnosti dostávat se do střeleckých pozic na ledové ploše, aniž by byly zohledněny výsledky střel. V grafu je vykreslený kumulativní součet hodnot.
- *Průměrný čas na střídání* – další metoda popisu role hráče. Delší doba střídání typicky značí větší zodpovědnost v rámci týmu. Krátká střídání zase svědčí o tom, že se hráč dostává na led spíše ve speciálních případech.

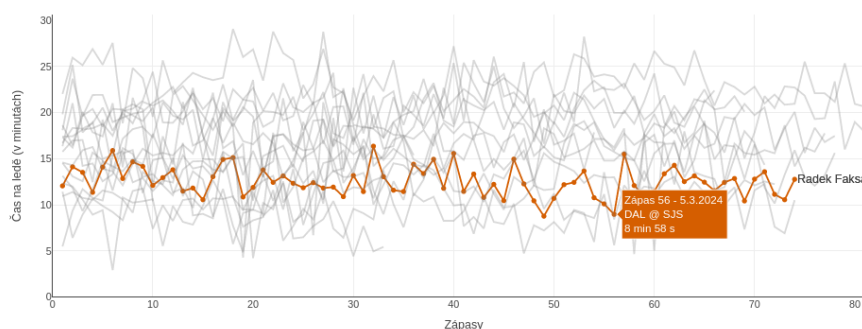
Výběr u skupiny brankářů poskytuje následující statistiky:

- *Čas na ledě* – u brankářů často odpovídá celé délce zápasu s lehkými výchyly. V případě prodloužení je čas delší než standardních 60 minut, pokud však tým prohrával ke konci zápasu,

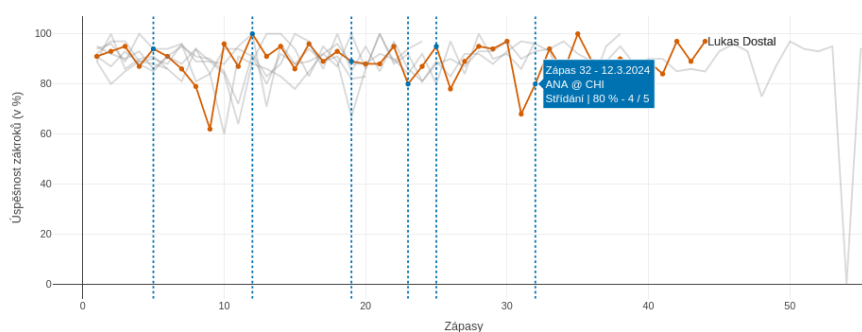
⁵Rostoucí čas na ledě poukazuje na to, že hráč podává dobré výkony. V opačném případě se může jednat naopak o změnu role v sestavě nebo ztrátu důvěry.

bývá zase kratší.⁶ Zajímavé jsou pak výraznější odchylky, které jsou v grafu vyjádřeny spolu s informací, zda v zápase nastoupil jako první brankář. Pokud má brankář krátký čas na ledě a v zápase nestadoval, pak zaskakoval za druhého brankáře, kterému se pravděpodobně nedařilo. Pokud však brankář v zápase má start, svědčí to o tom, že se jemu samotnému nedařilo a trenér se ho rozhodl odvolat.

- *Úspěšnost zákroků* – klíčová statistika pro brankáře. V grafu je popis jednotlivých bodů doplněn o údaje počtu střel, kterým brankář čelil. Ukázku lze vidět na obrázku 6.5.
- *Střely proti* – kumulativní součet počtu střel, kterým brankář během sezóny čelil.
- *Obdržené góly* – kumulativní součet obdržných gólů v sezóně.



■ **Obrázek 6.4** „Krátkodobé trendy“ pro bruslaře, zvolenou statistikou je *Čas na ledě* a hráčem Radek Faksa.



■ **Obrázek 6.5** „Krátkodobé trendy“ pro brankáře, zvolenou statistikou je *Úspěšnost zákroků*, zvoleným hráčem Lukáš Dostál. Modrou barvou jsou vyznačeny zápasy, ve kterých nestadoval.

6.3 Dlouhodobé trendy

Vizualizace s názvem „Dlouhodobé trendy“ má za cíl vystihnout vývoj role českých hráčů jako celku v NHL. Aby bylo možné takové údaje efektivně prezentovat, je potřeba nejprve shrnout data odpovídající jednotlivým hráčům tak, aby výsledná hodnota výstižně zachytila jejich výkony

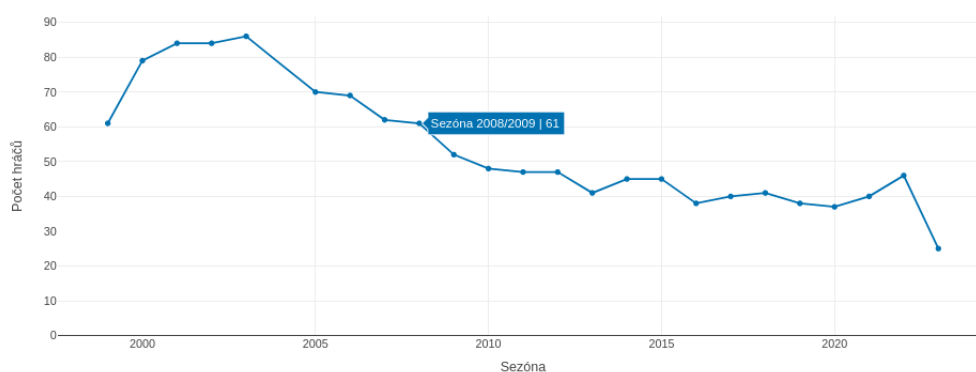
⁶To je dáno možností trenéra v jakoukoliv chvíli zápasu odvolat brankáře a nahradit ho dalším útočícím hráčem. Strategie s vysokým rizikem obdržení gólu se vyplácí zejména v situacích, kdy tým prohrává na konci základní hrací doby.

v rámci celé sezóny. Grafy v této sekci vychází z dat, jejichž stažení a zpracování je popsáno v podkapitole 5.2.4.

Jelikož podstatou grafu je sledovat vývoj vybraných proměnných v čase, byl pro vizualizaci zvolen *spojnicový graf*. Na ose x je v tomto případě vyneseno rok,⁷ oproti grafům v podkapitole 6.2, kde se jednalo o jednotlivé zápasy. Osa y pak odpovídá hodnotám zvolené statistiky.

Graf je definovaný dvojicí uživatelských vstupních proměnných. První z nich je výběr statistiky z následujících možností: *Počet hráčů*, *Čas na ledě*, *Góly* a *Body*. Jedná se o základní proměnné, které popisují role a výkony skupiny hráčů během sezóny. *Počet hráčů* odpovídá počtu všech českých hráčů, kteří se během dané sezóny objevili na soupisce některého z týmů, bez ohledu na to, kolik zápasů odehráli. Zároveň se jedná o jediný případ z nabídky statistik, kde jsou zahrnuti jak bruslaři, tak i brankáři. To je dáno tím, že statistiky *Góly* a *Body* popisují ryze útočnou fázi hry, na které se brankáři příliš často nepodílejí.⁸ U statistiky *Čas na ledě* pak platí podobné pravidlo, jako v podkapitole 6.2 – brankáři stráví na ledě nepoměrně více času za zápas, než ostatní hráči. Z tohoto důvodu nejsou brankářská data součástí ani této statistiky.

Údaj o počtu hráčů však nevystihuje míru zapojení jednotlivých hráčů v zápasech, jelikož zde jsou zahrnuti i hráči s nízkým časem na ledě nebo malým počtem odehraných zápasů. V tomto ohledu je výstižnější metrika o čase na ledě, která je v grafu zachycena dvěma způsoby – jako průměrná hodnota na jednoho hráče a jako sumární součet napříč všemi zápasy každého z hráčů. Stejným způsobem jsou rozděleny i zbývající dvě statistiky, které popisují počet vstřelených gólů a získaných bodů. Důvodem pro zahrnutí těchto metrik je vyjádření střelecké produktivity hráčů, spolu s podílem na akcích, které jsou ve výsledku nejdůležitější – vstřelených gólech. Druhým uživatelským vstupem je rozsah sezón k zobrazení. Kompletní přehled nabízí rozmezí od sezóny 1999/2000 až po aktuální, 2023/2024. Uživatel má však možnost vykreslené období libovolně omezit a zaměřit se jen na jeho část.



■ **Obrázek 6.6** „Dlouhodobé trendy“, zvolenou statistikou je *Počet hráčů*.

⁷Sezóny NHL typicky probíhají na přelomu dvou let. Kvůli přehlednosti grafu představuje každou sezónu první z dvojice let, kdy probíhala.

⁸Respektive i u brankářů jsou tyto statistiky vedeny, jelikož ve vzácných situacích si brankář asistenci připisuje. Ještě méně často se stává, že brankář skóruje gól (to nastává téměř výhradně v situaci, kdy soupeřící tým hraje bez brankáře). Z pohledu této práce jsou takové ojedinělé případy považovány za zanedbatelné, jelikož by brankářská data, v případě zahrnutí do výpočtu, měla velký vliv zejména na průměrnou hodnotu počtu gólů a asistencí na hráče.

Závěr

Tato práce se zabývá vizualizacemi statistik českých hokejistů v NHL. Nejprve byl proveden průzkum existujících webových stránek, které se zabývají prezentováním hokejových statistik zajímavým způsobem a nejen formou tabulkových záznamů. Přestože byla nalezena řada takových zdrojů, z nichž se většina zabývá hráči v NHL, nebyl objeven žádný, jehož zaměření by se podobalo této práci – vizualizace s důrazem na skupinu hráčů, odlišených podle národnosti. Následovala rešerše možných zdrojů dat, které poskytují potřebné údaje o hráčích. V této části bylo zjištěno, že jediným vhodným zdrojem, z hlediska obsáhlosti i spolehlivosti, je API poskytované přímo NHL. Přestože API nemá oficiální dokumentaci, podařilo se potřebná data získat za pomoci neoficiální dokumentace. Tento zdroj se ukázal být pro potřeby práce dostatečný, jelikož obsahuje nejen údaje o jednotlivých hráčích, ale také statistické souhrny jednotlivých zápasů v sezóně a data týkající se týmů – pořadí týmů v lize, plán zápasů a především soupisky týmů podle sezóny. Avšak ukázalo se, že ani tento zdroj není naprosto stabilní a spolehlivý. Před začátkem sezóny 2023/2024 došlo ke změně struktury celého API a tudíž změně URL adres jednotlivých endpointů. Původní adresy nebyly zachovány a bylo tak třeba aktualizovat veškeré skripty, které slouží ke stažení a zpracování dat. V případě další podobné změny API by bylo nutné udělat podobnou úpravu skriptů znovu.

V praktické části práce byla nejprve navržena a později vytvořena sada vizualizací a grafů, jejichž cílem bylo prezentovat statistiky českých hráčů ze tří pohledů – popis výkonů v jednotlivých zápasech, přehled vybraných měřítek v rámci sezóny a vývoj role českých hokejistů jako celku v NHL. Motivací pro zvolení tématu bakalářské práce bylo poskytnout hokejovým fanouškům přehled o aktuálním dění v NHL z pohledu českých hokejistů, což bylo reflektováno během celého procesu návrhu vizualizací. Výsledná sada grafů byla rozdělena do tří sekcí odpovídajícím způsobem.

V sekci „Dnešní dění“ jsou shrnuty veškeré důležité akce zvoleného hráče během jednoho zápasu, který se odehrál ve vybraný den. Dva z grafů v této sekci prezentují místa, kde se na ledové ploše odehrály akce spojené s tímto hráčem. Z umístění bodů lze pak odvodit dodatečný kontext, který blíže vystihuje jednotlivé akce. Prvním takovým grafem je *Schéma střel*. Jedná se o *bodový graf*, kde jsou vykresleny veškeré střely českých hráčů. Jednotlivé střely jsou barevně odlišeny na čtyři kategorie – střely na branku, mimo branku, zblokované střely a góly. Z jejich umístění na hrací ploše pak lze odvodit, jak byly jednotlivé střely nebezpečné pro soupeře. Pokud se střela nachází v těsné blízkosti branky soupeřícího týmu, vypovídá to o vysoké šanci skórovat gól. Naopak střely, které se nachází daleko od branky, bývají méně nebezpečné. Umístěním zápasových akcí se zabývá také *Schéma akcí*. Jedná se opět o *bodový graf*, ale tentokrát jsou vykresleny všechny ostatní zápasové akce, kromě střel. Jejich umístění na hrací ploše znovu vyjadřuje okolnosti dané akce, podobně jako u předchozího schématu. Posledním grafem v sekci je *Čas na ledě*, jehož cílem je prezentovat hráče, se kterými český hráč strávil nejvíce času na ledě. Z této informace lze odhadnout roli hráče v týmu. Obecně platí, že kvalita spoluhráčů,

se kterými se hráč na ledě nejčastěji potkával, do jisté míry vypovídá i o úrovni hry zvoleného hráče. Podobné pravidlo platí i ohledně soupeřících hráčů.

Kategorie „Krátkodobé trendy“ se zabývá prezentací statistik hráčů ve všech zápasech aktuální sezóny. Jelikož se značně liší měřítka pro vyjádření výkonu bruslařů a brankářů, byly vytvořeny dva odlišné grafy podle hráčské pozice. Každý z grafů vystihuje vývoj vybraných metrik pro skupinu hráčů. Z grafu týkajícího se bruslařů lze zjistit, jak se vyvíjí role hráče v týmu, podle údaje o času na ledě v jednotlivých zápasech. Dále graf popisuje střeleckou produktivitu hráče během sezóny. Hodnoty vybraných metrik – počet vstřelených gólů, získaných bodů a střel – lze také porovnat s ostatními českými hráči v sezóně. Výkon brankářů pak vystihuje statistika procentuální úspěšnosti zákroků v zápasech, což je pro brankáře klíčové měřítko. Dále je možné vzájemně porovnat české brankáře v počtu střel, kterým čelili za sezónu, nebo počtu obdržených gólů.

Cílem sekce „Dlouhodobé trendy“ byla prezentace vybraných statistik v měřítku několika sezón pro české hráče v NHL. Záměrem bylo vystihnout vývoj těchto metrik v dlouhodobém rozmezí a popsat tak změnu role českých hráčů jako celku. Klíčová je metrika *Počet hráčů*, kde jsou zahrnuti všichni hráči, kteří během sezóny nastoupili za některý z týmů. Údaj však nevyovídá o tom, jak aktivně se hráči skutečně zapojili do hry. Za účelem vyjádřit střeleckou produktivitu hráčů byly zvoleny statistiky *Góly* a *Body*, *Čas na ledě* pak popisuje míru zapojení hráčů do zápasů. Zmíněné statistiky jsou vždy prezentovány ve dvou variantách – v průměru na jednoho hráče a jako celkový součet hodnot všech hráčů. Z grafů lze zjistit, že ačkoliv celkový počet českých hráčů v NHL klesá, jejich kvalita zůstává velmi dobrá. Dokonce se ukazuje, že na základě průměrného počtu gólů a asistencí si vedou lépe, než v předchozích letech. Trend vypovídá o tom, že ačkoliv NHL v poslední době opustilo mnoho Čechů, pravděpodobně se z velké části jednalo o hráče s nižší výkonností. V soutěži pak zůstali převážně nejvíce produktivní hráči. Výsledný graf však opomíjí statistiky brankářů. Vyjmutí brankářských statistik z grafů odpovídajících počtu vstřelených gólů a získaných bodů je opodstatněné. Zahrnutím těchto údajů by se hodnoty v podobě sumárního součtu změnily jen velmi nepatrně. Průměrných hodnoty těchto statistik na hráče by pak brankářské údaje zkreslovaly nežádoucím způsobem – brankáři si připisují minimum gólů a bodů celkově a výsledná průměrná hodnota by byla nižší. Podobným způsobem by zahrnutí brankářů nepoměrně ovlivnilo i statistiku *Čas na ledě*. V tomto případě by naopak docházelo k nepoměrnému zvýšení hodnot – brankáři mívají nejvyšší čas na ledě, často i několikanásobně vyšší než bruslaři.

Na závěr byl vytvořený prototyp webové aplikace pro prezentaci vybraných grafů. Aplikace je vytvořena v programovacím jazyce R s využitím balíčku *shiny*, který poskytuje rámec pro vývoj webových aplikací zaměřených na prezentaci grafů a datovou analýzu. Součástí jsou i skripty v jazyce R, které data stahují z API a ukládají v upravené podobě do lokální instance databáze MongoDB v rámci Docker kontejneru. Jak jednotlivé skripty, tak i aplikaci samotnou je možné spustit lokálně v rámci stejného kontejneru. Skripty je možné spouštět nezávisle na sobě za účelem aktualizace existujících dat v databázi.

Bibliografie

1. JAMES, B. *The Bill James Historical Baseball Abstract*. Villard Books, 1986. ISBN 9780394537139. Dostupné také z: <https://books.google.cz/books?id=-JcYAAAAIAAJ>.
2. *Moneyball – Summary of 2 key ideas* [<https://www.blinkist.com/en/books/moneyball-en>]. [B.r.]. [Accessed 03-04-2024].
3. *The New Bill James Historical Baseball Abstract — enbook.cz* [<https://www.enbook.cz/the-new-bill-james-historical-baseball-abstract>]. [B.r.]. [Accessed 03-04-2024].
4. LEWIS, M. *Moneyball: The Art Of Winning An Unfair Game*. WW Norton, 2003. Business book summary. ISBN 9780393057652. Dostupné také z: https://books.google.cz/books?id=RWOX_2eYPcAC.
5. *Moneyball: The Art of Winning an Unfair Game — barnesandnoble.com* [<https://www.barnesandnoble.com/w/moneyball-michael-lewis/1005525269>]. [B.r.]. [Accessed 03-04-2024].
6. ISAAC, Dave. *NHL introduces puck- and player-tracking technology — eu.usatoday.com* [<https://eu.usatoday.com/story/sports/nhl/allstar/2015/01/25/nhl-player-puck-tracking-technology/22315241/>]. 2015. [Accessed 04-04-2024].
7. *NHL EDGE Puck and Player Tracking Statistics* [<https://edge.nhl.com/en/home>]. [B.r.]. [Accessed 15-03-2024].
8. WYSHYNSKI, Greg. *NHL brings advanced puck tracking stats to public — espn.com* [https://www.espn.com/nhl/story/_/id/38724195/nhl-brings-advanced-puck-tracking-stats-public]. 2023. [Accessed 04-04-2024].
9. ENGEL, Andrew. *Why Sports Analytics Matters — datarobot.com* [<https://www.datarobot.com/blog/why-sports-analytics-matters/>]. 2019. [Accessed 12-05-2024].
10. *History of Ice hockey in Timeline - Popular Timelines — populartimelines.com* [<https://populartimelines.com/timeline/Ice-hockey>]. [B.r.]. [Accessed 15-03-2024].
11. ESKENAZI, Gerald; FISCHLER, Stanley I.; FISCHLER, Shirley W. *Ice hockey | History, Rules, Equipment, Players, & Facts — britannica.com* [<https://www.britannica.com/sports/ice-hockey>]. 2024. [Accessed 15-03-2024].
12. BENJAMIN, Daniel. *Traditions Pro Vs. Amateur - TIME — web.archive.org* [<https://web.archive.org/web/20090902183140/http://www.time.com/time/magazine/article/0,9171,976117-1,00.html>]. 1992. [Accessed 25-03-2024].
13. *A Brief History of the NHL — nhl.com* [<https://www.nhl.com/history/a-brief-history-of-the-league>]. [B.r.]. [Accessed 15-03-2024].

14. *Rules of Hockey for Dummies: Hockey Beginner's guide* — owayo.com [<https://www.owayo.com/magazine/rules-of-hockey-made-simple-us.htm>]. [B.r.]. [Accessed 15-03-2024].
15. *NHL.CZ* [<https://nhl.cz/>]. [B.r.]. [Accessed 15-03-2024].
16. OKABAYASHI, Sai; IVERSON, Erik. *ShiftChart - Hockey Shift Data Visualization* — shiftchart.com [<http://shiftchart.com/>]. [B.r.]. [Accessed 12-05-2024].
17. MCCURDY, Micah Blake. *HockeyViz* — hockeyviz.com [<https://hockeyviz.com>]. [B.r.]. [Accessed 15-03-2024].
18. *NHL Players by Country 2024* — worldpopulationreview.com [<https://worldpopulationreview.com/country-rankings/nhl-players-by-country>]. [B.r.]. [Accessed 15-03-2024].
19. *The GNU General Public License v3.0 - GNU Project - Free Software Foundation* — gnu.org [<https://www.gnu.org/licenses/gpl-3.0.html>]. 2023. [Accessed 17-04-2024].
20. *R: What is R?* — r-project.org [<https://www.r-project.org/about.html>]. [B.r.]. [Accessed 15-04-2024].
21. GIORGI, Federico M.; CERAOLO, Carmine; MERCATELLI, Daniele. *The R Language: An Engine for Bioinformatics and Data Science* — ncbi.nlm.nih.gov [<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9148156/>]. [B.r.]. [Accessed 15-04-2024].
22. *The Comprehensive R Archive Network* — cran.r-project.org [<https://cran.r-project.org/>]. [B.r.]. [Accessed 28-04-2024].
23. *Download RStudio | The Popular Open-Source IDE from Posit* — posit.co [<https://posit.co/products/open-source/rstudio/>]. [B.r.]. [Accessed 15-04-2024].
24. SCHAEFER, Lauren. *What Is NoSQL? NoSQL Databases Explained* — mongodb.com [<https://www.mongodb.com/nosql-explained>]. [B.r.]. [Accessed 15-04-2024].
25. TEJADA, Zoiner. *Non-relational data and NoSQL - Azure Architecture Center* — learn.microsoft.com [<https://learn.microsoft.com/en-us/azure/architecture/data-guide/big-data/non-relational-data>]. [B.r.]. [Accessed 17-04-2024].
26. *What is a Container?* — docker.com [<https://www.docker.com/resources/what-container/>]. [B.r.]. [Accessed 29-04-2024].
27. VIVEK, Ratan. *Docker: A Favourite in the DevOps World - open source for you* — opensourceforu.com [<https://www.opensourceforu.com/2017/02/docker-favourite-devops-world/>]. 2017. [Accessed 29-04-2024].

Obsah přiloženého média

	readme.txt	stručný popis obsahu média
	impl.	zdrojové kódy implementace
	text	text práce
		thesis zdrojová forma práce ve formátu L ^A T _E X
		thesis.pdf text práce ve formátu PDF