

Projekt 2. část - dokumentace výsledků

Varianta 3: COVID-19

Ukládání a příprava dat

Šimon Galba, Bc. (xgalba03)
Gladiš Damián, Bc. (xgladi00)
Jeřábek František, Bc. (xjerab25)

16. prosince 2021

Obsah

1	Dotaz A - 1	1
2	Dotaz A - 2	2
3	Dotaz B - 1	2
4	Vlastne dotazy	4
4.1	Dotaz 1	4
4.2	Dotaz 2	5

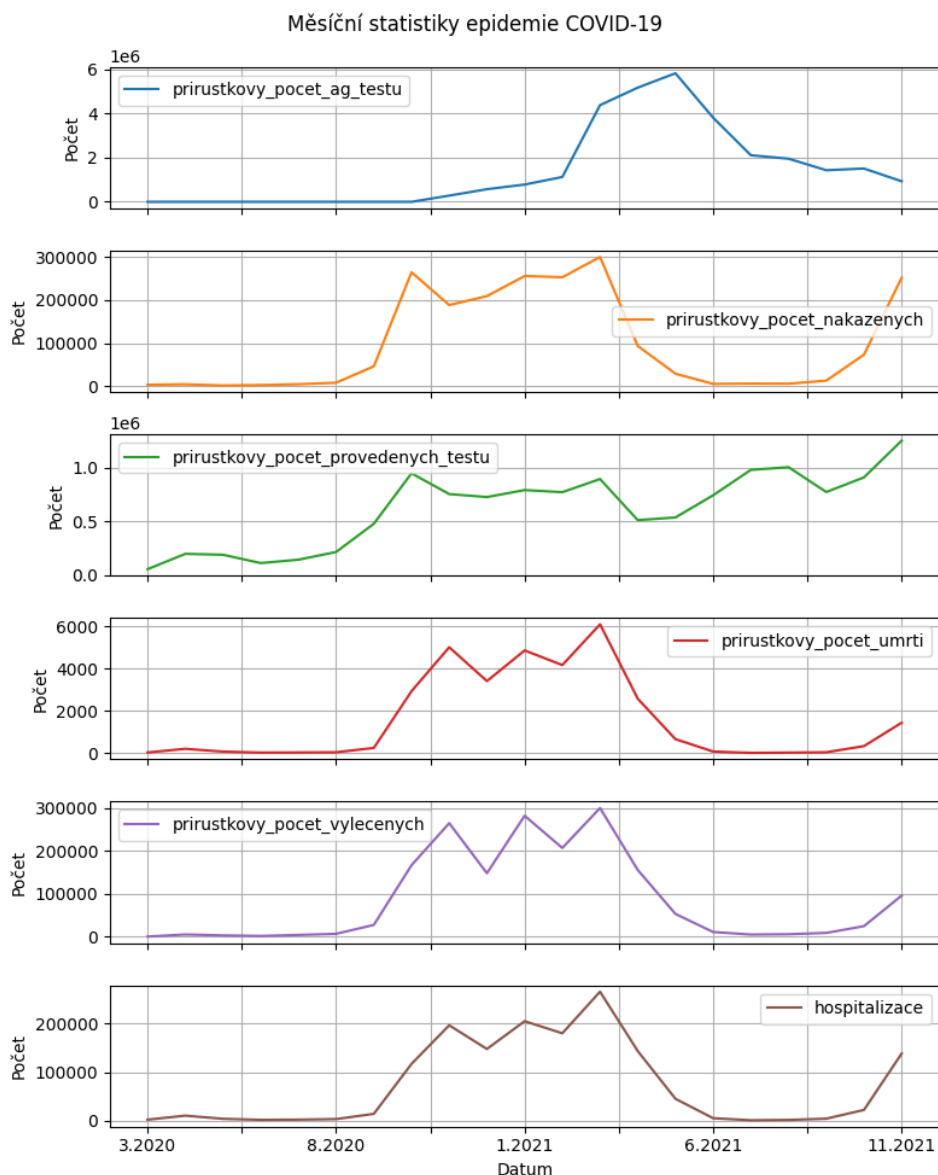
1 Dotaz A - 1

Vytvořte čárový (spojnicový) graf zobrazující vývoj covidové situace po měsících pomocí následujících hodnot: počet nově nakažených za měsíc, počet nově vyléčených za měsíc, počet nově hospitalizovaných osob za měsíc, počet provedených testů za měsíc. Pokud nebude výsledný graf dobře čitelný, zvažte logaritmické měřítko, nebo rozdělte hodnoty do více grafů.

Výsledky:

Pomocou agregační funkce `aggregate` boli data získané pomocou dotazov **group**, **project** a **sort** priamo z databáze MongoDB vytvorenej v prvej časti projektu. Pre získanie mesačných dát bolo použité dotazovacie ID ako kombinácia mesiac-rok, data boli následne sčítané pre počet testov, počet infikovaných, vyliečených a počet úmrtí.

Pre lepšiu prehľadnosť bolo vytvorených 5 grafov zobrazujúcich jednotlivé hodnoty opisujúce pandemickú situáciu v Českej Republike. Počty testov (graf 1 a 3) sú uvedené v **miliónoch**.



Obrázek 1: Grafy zobrazující vývoj pandemickej situácie v ČR po mesiacoch.

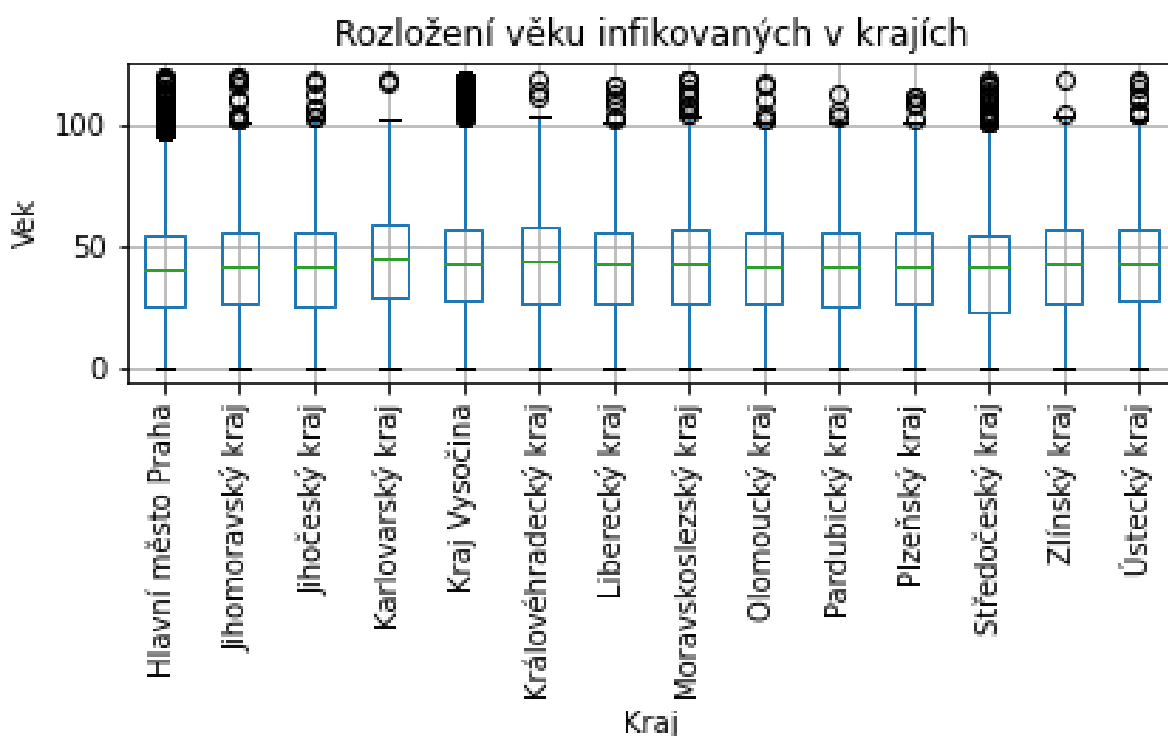
2 Dotaz A - 2

Vytvořte krabicové grafy zobrazující rozložení věku nakažených osob v jednotlivých krajích.

Výsledky:

Pomocou agregační funkce **aggregate** boli dáta získané pomocou dotazov **match**, **group**, **project** a **sort** priamo z databáze mongoDB vytvorenej v prvej časti projektu. Pre získanie dát krajov bolo použité dotazovacie ID ako kombinácia kód kraju-vek. Z databáze boli odstránené dáta ktoré neobsahovali hodnotu vek alebo kód kraju.

Z grafu sa dá vyčítať, že jednotlivé kraje nemali vplyv na priemerný vek očkovaných, medián jednotlivých krajov sa nachádza na zhruba rovnakej hodnote 45 rokov.



Obrázek 2: Graf zobrazující rozložení věku infikovaných v jednotlivých krajích ČR.

3 Dotaz B - 1

Sestavte 4 žebříčky krajů "best in covid" za poslední 4 čtvrtletí (1 čtvrtletí = 1 žebříček). Jako kritérium volte počet nově nakažených přepočtený na jednoho obyvatele kraje. Pro jedno čtvrtletí zobrazte výsledky také graficky. Graf bude pro každý kraj zobrazovat celkový počet nově nakažených, celkový počet obyvatel a počet nakažených na jednoho obyvatele. Graf můžete zhotovit kombinací dvou grafů do jednoho (jeden sloupcový graf zobrazí první dvě hodnoty a druhý, čárový graf, hodnotu třetí).

Výsledky:

Pomocou funkce **aggregate** a dotazov **match**, **group** a **project** sú opäť priamo z databáze mongoDB vytvorenej v prvej časti projektu získané dáta o infikovaných osobách v jednotlivých krajoch. Následné sú tieto dáta združené pomocou funkcie **groupby** a parametrov **rok**, **štvrtrok** a **kraj** a zoradené funkcou **sort** pre vytvorenie

rebríčka "best in covid" pre jednotlivé štvrťroky od posledného v roku 2020 až po tretí v roku 2021. Z dát môžeme vyčítať, že v zvolených štvrťrokoch sa situácia postupne zlepšovala a aj, že žiadny kraj dlhodobo neobsadil poprednú priečku v rebríčku.

Tabulka 1: 4. čtvrtletí 2020

	Název kraje	infected/person
1	Zlínský kraj	0.080365
2	Kraj Vysočina	0.071610
3	Královéhradecký kraj	0.071029
4	Liberecký kraj	0.068562
5	Pardubický kraj	0.067114
6	Moravskoslezský kraj	0.066747
7	Olomoucký kraj	0.065252
8	Jihočeský kraj	0.062719
9	Středočeský kraj	0.061444
10	Plzeňský kraj	0.059307
11	Ústecký kraj	0.057021
12	Jihomoravský kraj	0.052980
13	Hlavní město Praha	0.050426
14	Karlovarský kraj	0.046275

Tabulka 2: 1. čtvrtletí 2021

	Název kraje	infected/person
1	Královéhradecký kraj	0.109952
2	Karlovarský kraj	0.099093
3	Liberecký kraj	0.097101
4	Pardubický kraj	0.094359
5	Plzeňský kraj	0.092882
6	Středočeský kraj	0.086301
7	Ústecký kraj	0.075729
8	Jihočeský kraj	0.072428
9	Hlavní město Praha	0.067423
10	Kraj Vysočina	0.064514
11	Olomoucký kraj	0.063485
12	Moravskoslezský kraj	0.062369
13	Jihomoravský kraj	0.060759
14	Zlínský kraj	0.054725

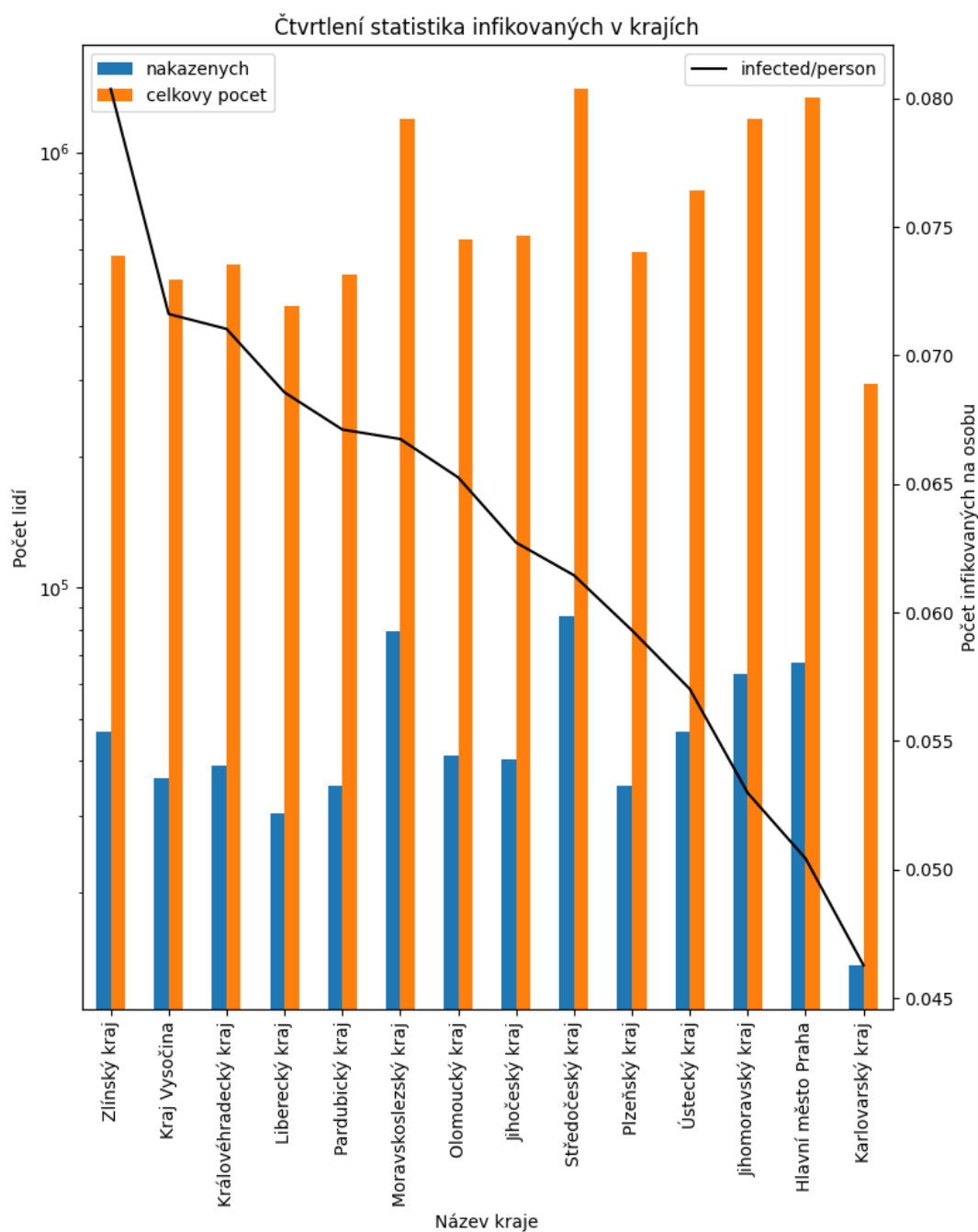
Tabulka 3: 2. čtvrtletí 2020

	Název kraje	infected/person
1	Zlínský kraj	0.017017
2	Jihočeský kraj	0.016157
3	Kraj Vysočina	0.015702
4	Ústecký kraj	0.014658
5	Moravskoslezský kraj	0.014222
6	Olomoucký kraj	0.013652
7	Pardubický kraj	0.012120
8	Liberecký kraj	0.011718
9	Jihomoravský kraj	0.011652
10	Středočeský kraj	0.010310
11	Hlavní město Praha	0.009721
12	Plzeňský kraj	0.008133
13	Královéhradecký kraj	0.004995
14	Karlovarský kraj	0.003870

Tabulka 4: 4. čtvrtletí 2021

	Název kraje	infected/person
1	Hlavní město Praha	0.004667
2	Plzeňský kraj	0.002775
3	Středočeský kraj	0.002718
4	Jihočeský kraj	0.002238
5	Moravskoslezský kraj	0.002198
6	Karlovarský kraj	0.002073
7	Jihomoravský kraj	0.001961
8	Pardubický kraj	0.001698
9	Kraj Vysočina	0.001629
10	Liberecký kraj	0.001503
11	Zlínský kraj	0.001501
12	Ústecký kraj	0.001448
13	Olomoucký kraj	0.001337
14	Královéhradecký kraj	0.001068

Pre grafické znázornenie bol vybraný 4. štvrťrok roku 2020. Tento graf je oproti tabuľkám doplnený o celkové počty infikovaných a o celkový počet obyvateľov kraja. Potom bol funkciou **sort** zoradený podľa počtu infikovaných na jednu osobu (čiarový graf).

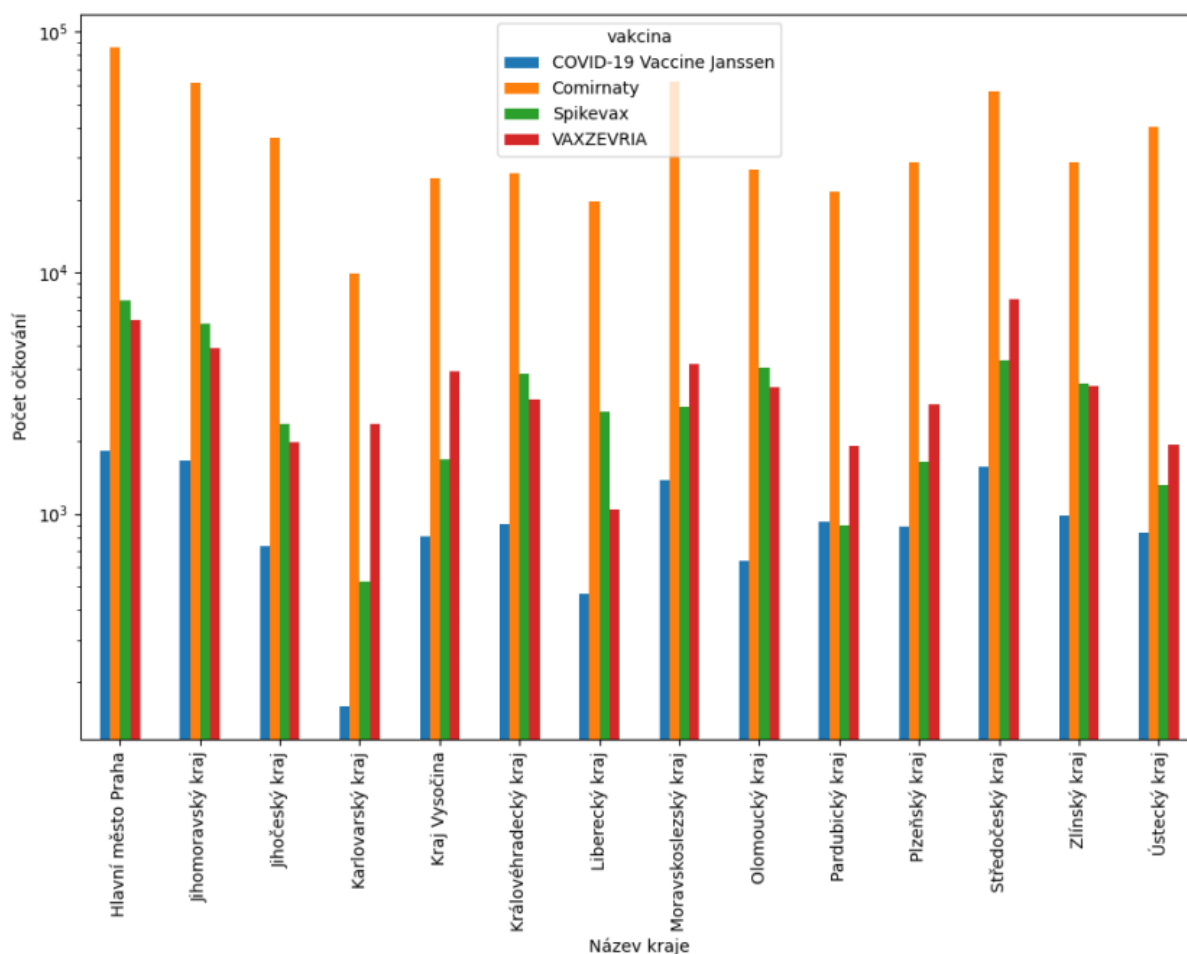


Obrázek 3: Graficky zobrazená statistika pre 4 štvrťrok roku 2020. Graf je zoradený podľa prírastku infikovaných osôb na jedného obyvateľa v prepočte na počet obyvateľov.

4 Vlastne dotazy

4.1 Dotaz 1

Pre prvý vlastný dotaz sme z databázy vytvorili prehľadný graf použitých vakcín pre jednotlivé kraje ČR. Dáta boli získané z datasetu očkovaných ľudí ČR uloženého v databáze. Dáta boli priamo v databáze agregované a pomocou dotazov **group** a **project** boli získané potrebné dáta vo formáte csv, ktoré boli následne prenesené do grafu.

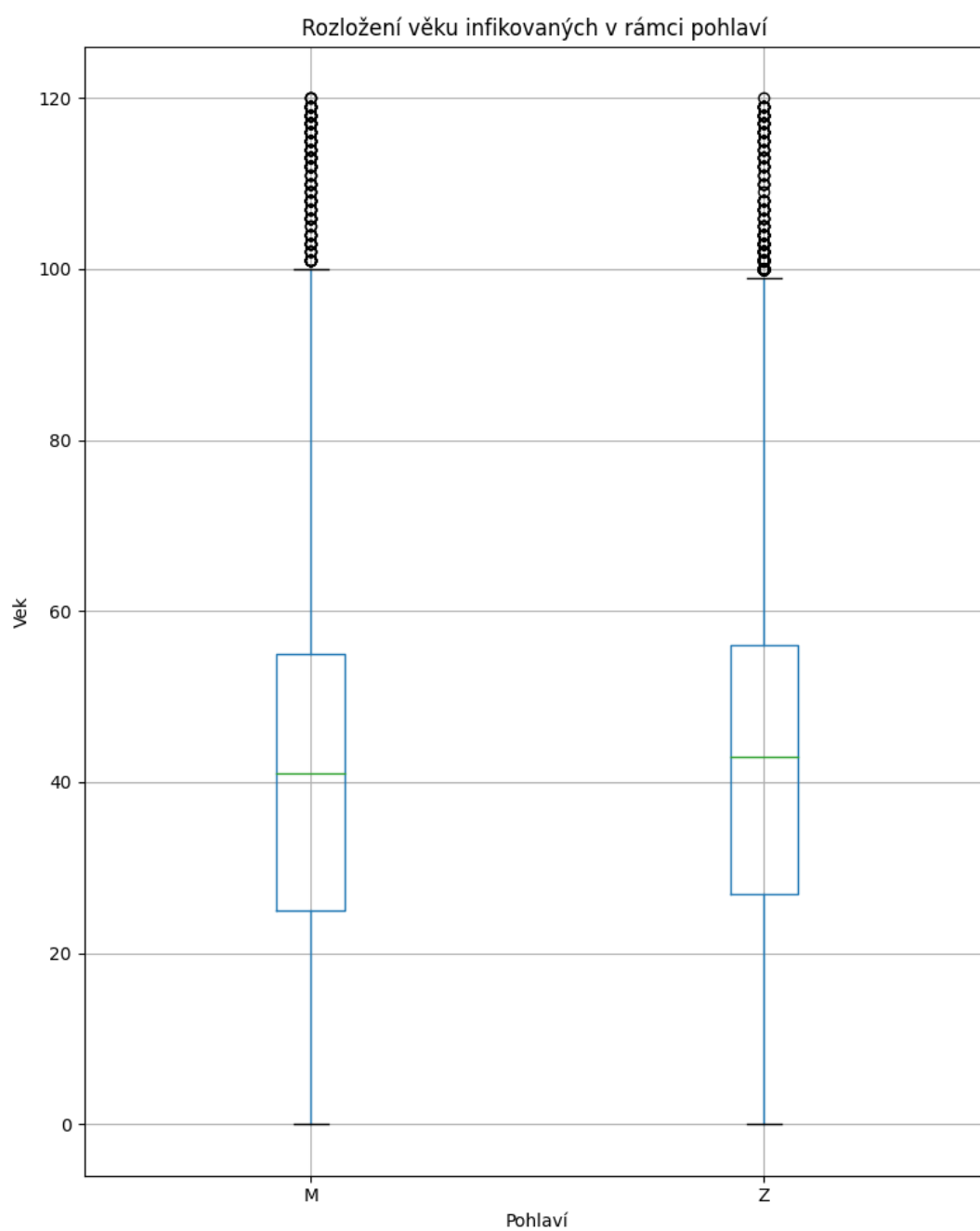


Obrázek 4: Rozdelenie použitých vakcín v jednotlivých krajoch

4.2 Dotaz 2

Pre druhý vlastný dotaz sme sa rozhodli zistiť priemerný počet nakazených mužov a žien a tieto hodnoty graficky porovnať. Pre znázornenie sme sa rozhodli použiť krabicový graf a z neho je možné vyčítať, že priemerný vek nakazených žien je vyšší ako priemerný vek nakazených mužov.

Výsledky boli získané pomocou agregácie datasetu infikovaných v jednotlivých krajoch spojenom pre celú republiku dotazmi **match**, **group** a **project** a následnom sčítaní podľa pohlavia a veku. Boli odfiltrované dáta neobsahujúce vek pacienta (čistenie dát).



Obrázek 5: Krabicový graf věku infikovaných osob v závislosti na pohlaví