Univerzita Pardubice
Fakulta elektrotechniky a informatiky

Zpracování dat pro předmět NMAST

Bc. Lukáš Milar, Bc. Tomáš Prudký

Semestrální práce

2021

OBSAH

SEZNAM OBRÁZKŮ

SEZNAM TABULEK

ÚVOD

1 POPIS DAT

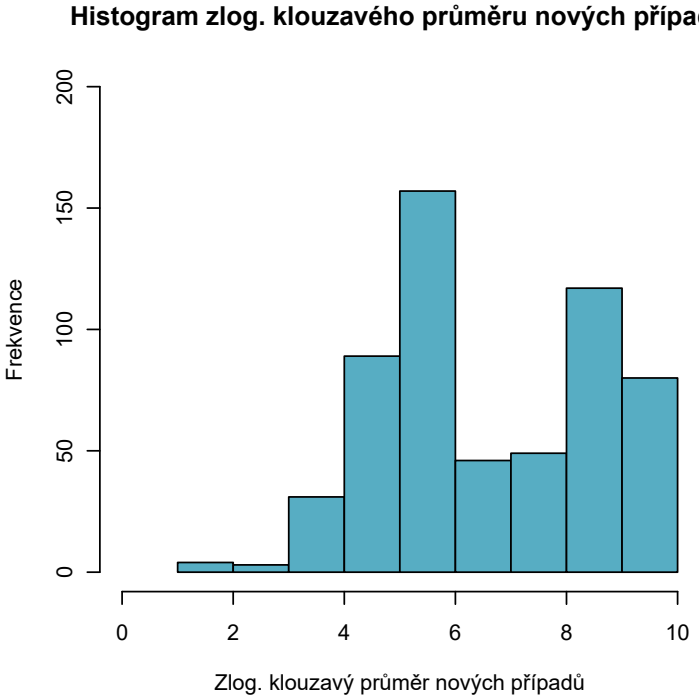
2 POPISNÁ STATISTIKA

	n_p	n_p_s	n_p_n_m	n_p_n_m_s	h_p	h_p_n_m
prumer	2940.58	2936.89	274.19	273.85	2370.49	221.03
modus	75.00	57.57	6.99	5.37	69.00	6.43
median	416.00	422.29	38.79	39.38	339.00	31.61
max	17773.00	12954.86	1657.22	1207.96	9509.00	886.66
min	-2214.00	2.71	-206.44	0.25	0.00	0.00
skewness	1.55	1.16	1.55	1.16	0.86	0.86
kurtosis	1.38	-0.07	1.38	-0.07	-0.86	-0.86
deviation	4277.10	3876.20	398.81	361.43	2973.01	277.22
var	18293577.15	15024928.55	159052.40	130633.34	8838793.84	76848.36

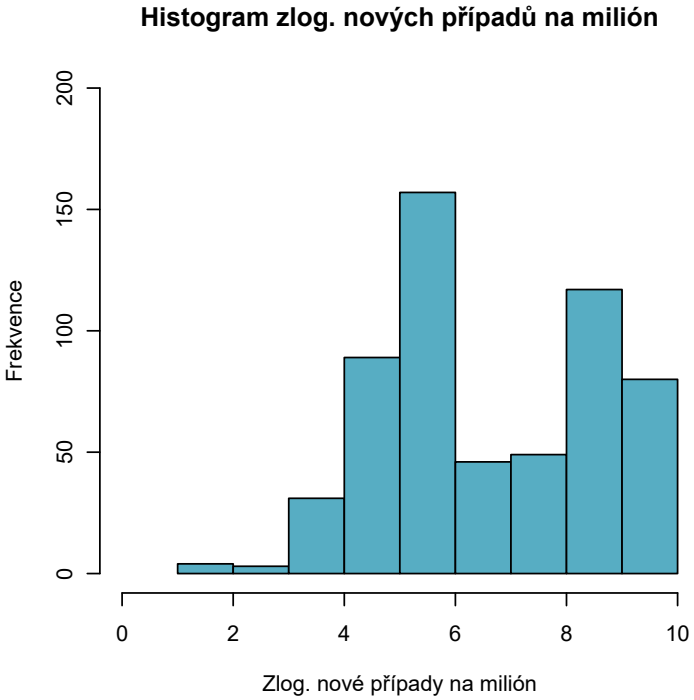
Tabulka 1: Části popisné statistiky aplikované na data

3 ZÁKLADNÍ GRAFY

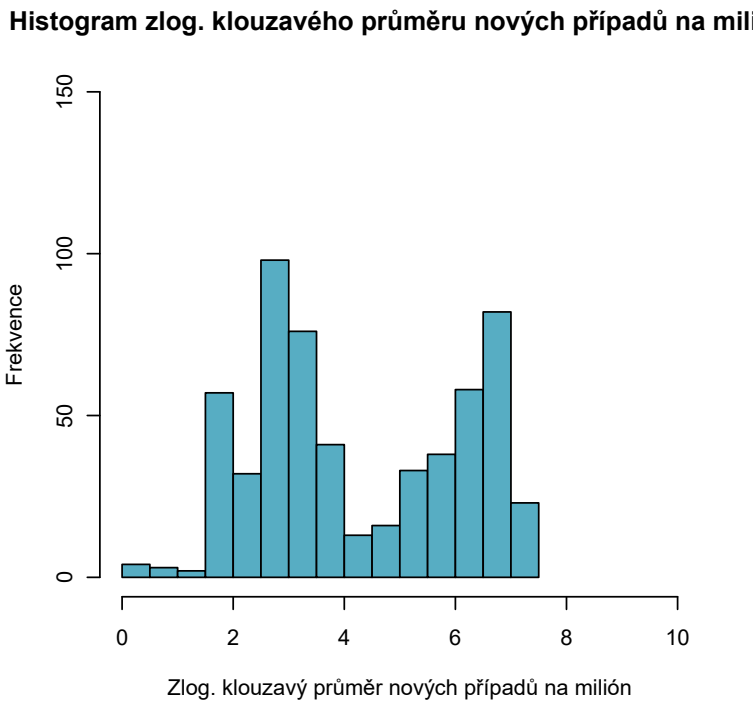
3.1 Histogram



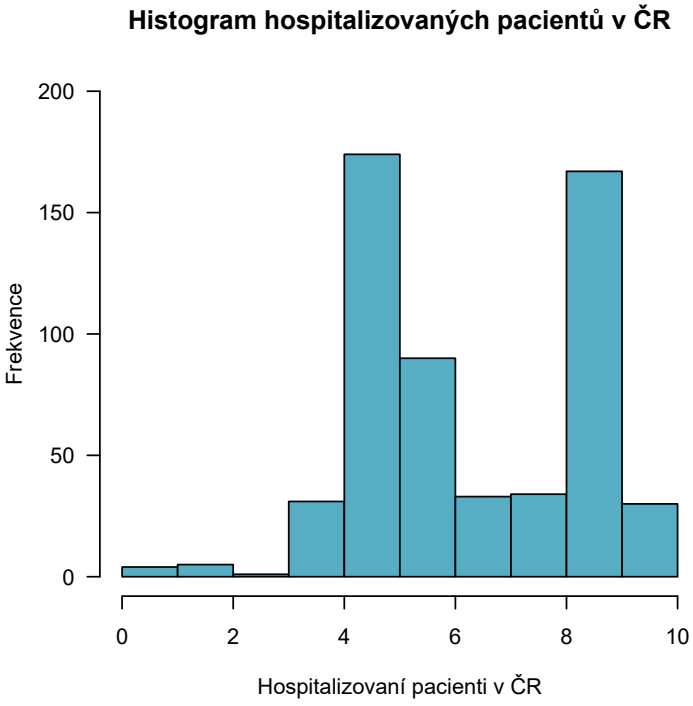
Obrázek 1: Histogram zlog. klouzavý průměr nových případů



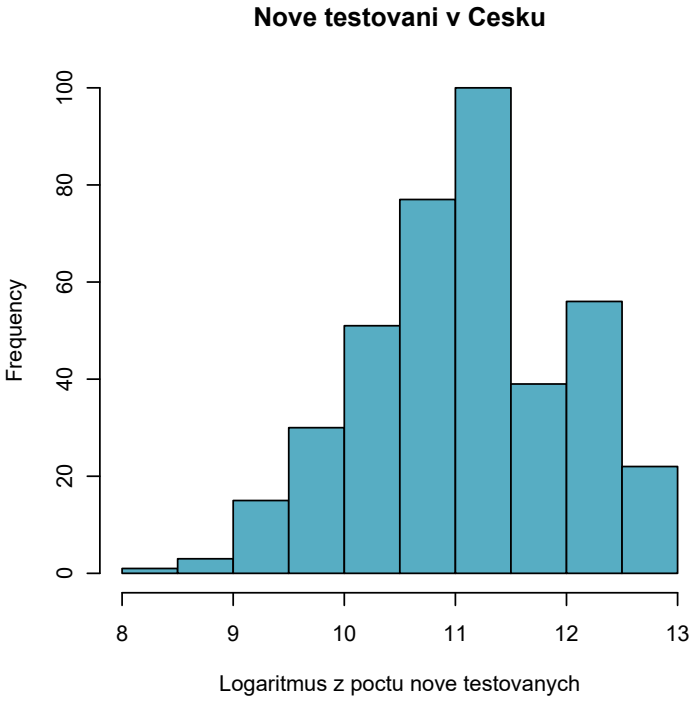
Obrázek 2: Histogram zlog. nových případů na milión



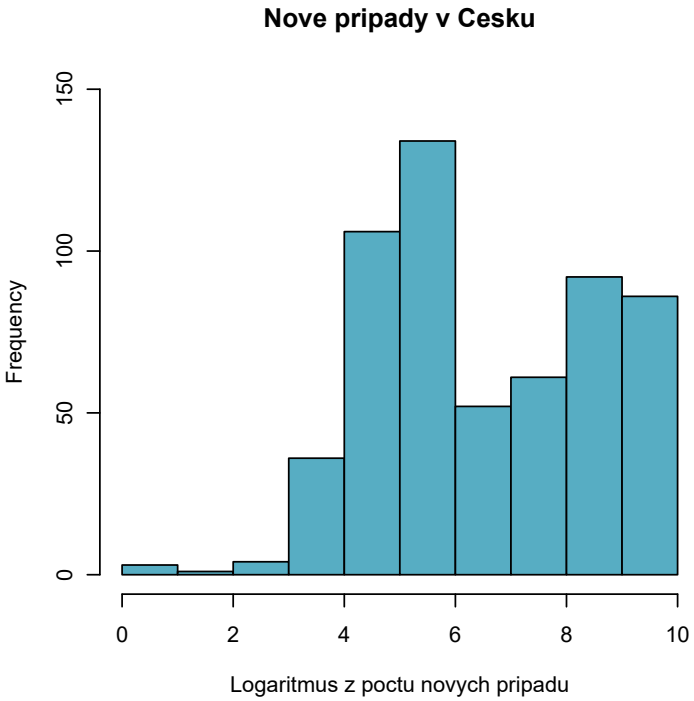
Obrázek 3: Histogram zlog. klouzavého průměru nových případů na milión



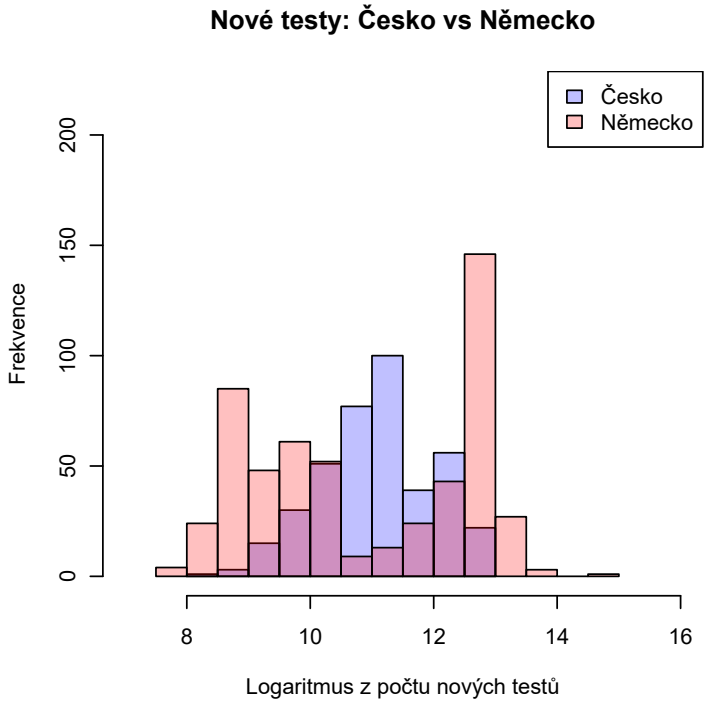
Obrázek 4: Histogram hospitalizovaných pacientů v ČR



Obrázek 5: Nove testovani v Cesku

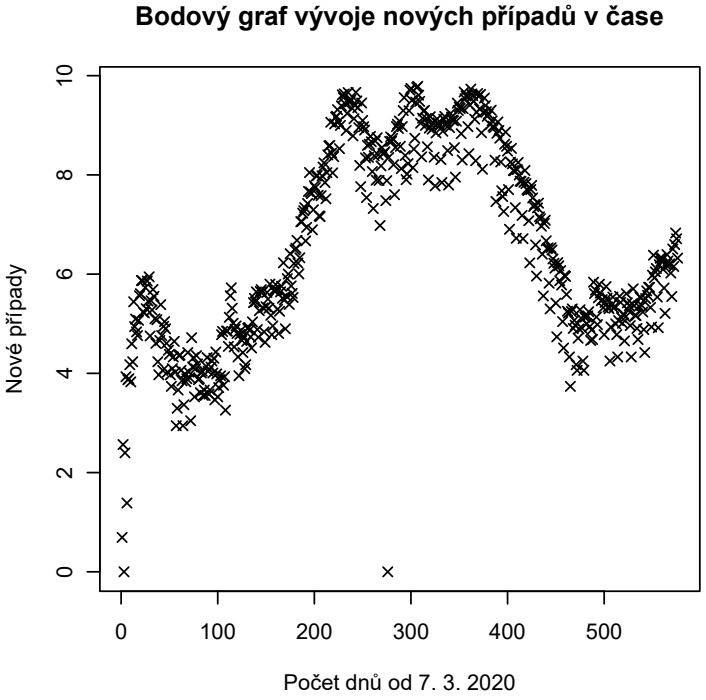


Obrázek 6: Nove pripady v Cesku

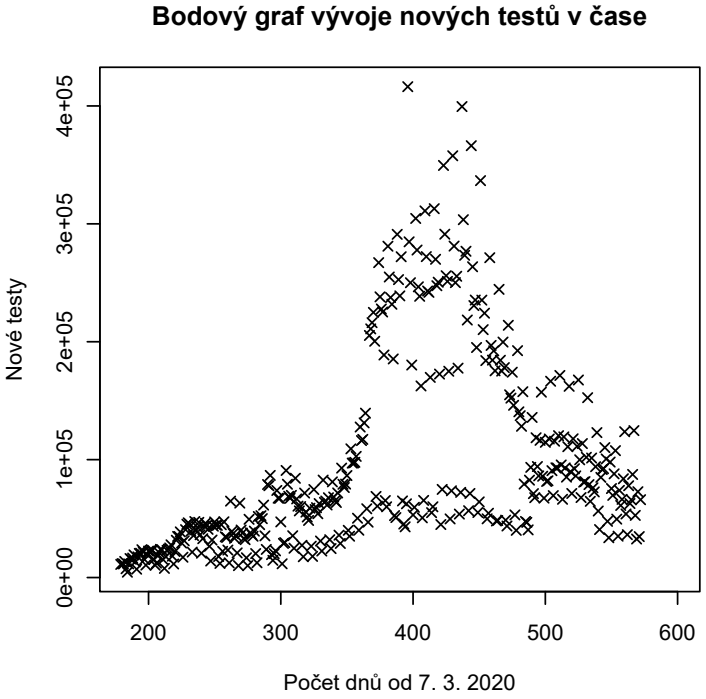


Obrázek 7: Nové testy pro Česko a Německo

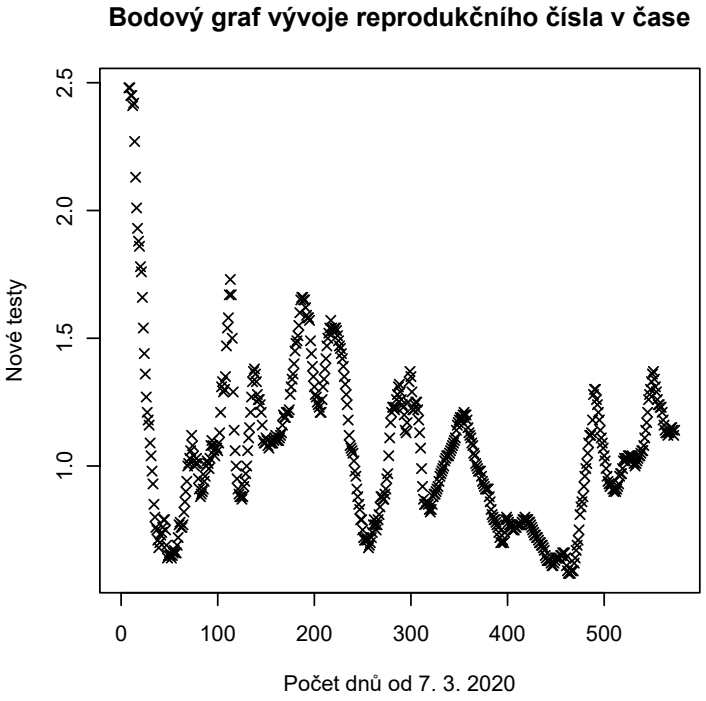
3.2 Bodový graf



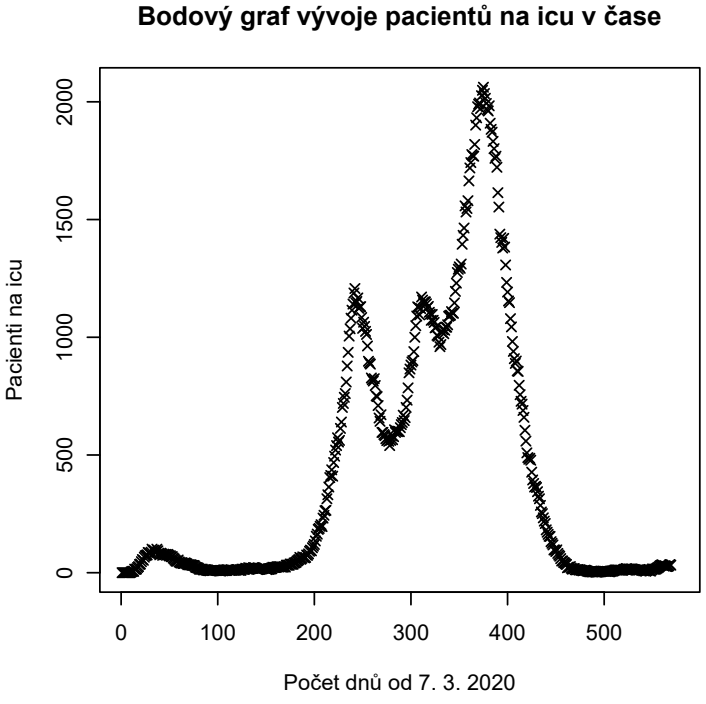
Obrázek 8: Bodový graf zlogaritmovaných nových případů



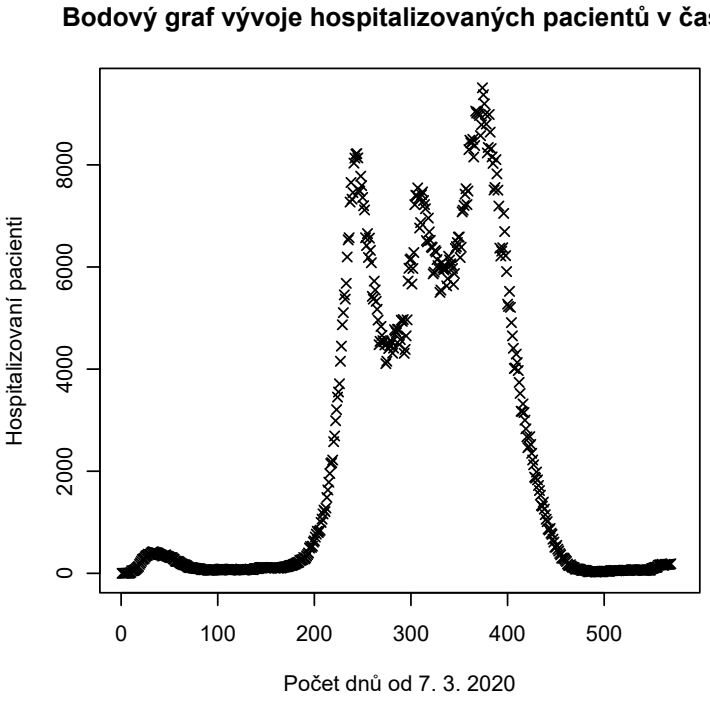
Obrázek 9: Bodový graf nových testů



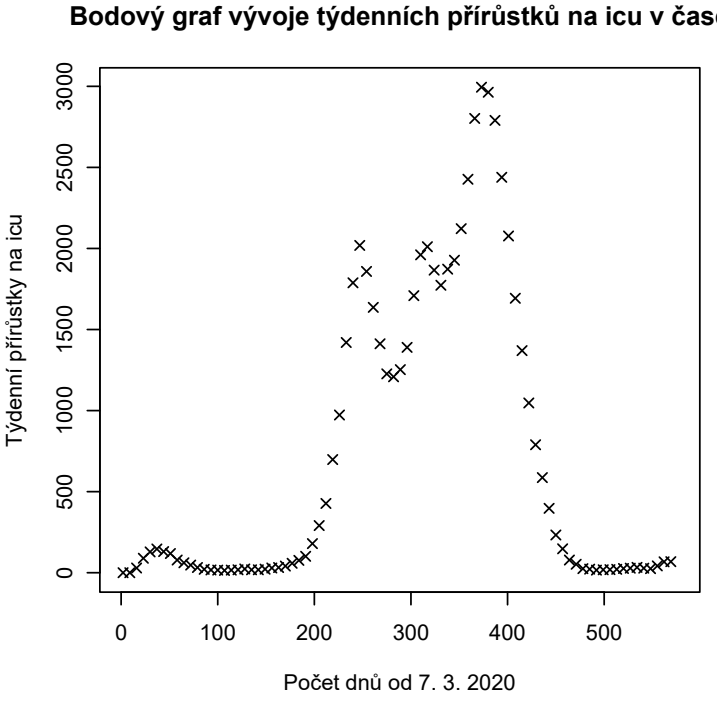
Obrázek 10: Bodový graf reprodukčního čísla



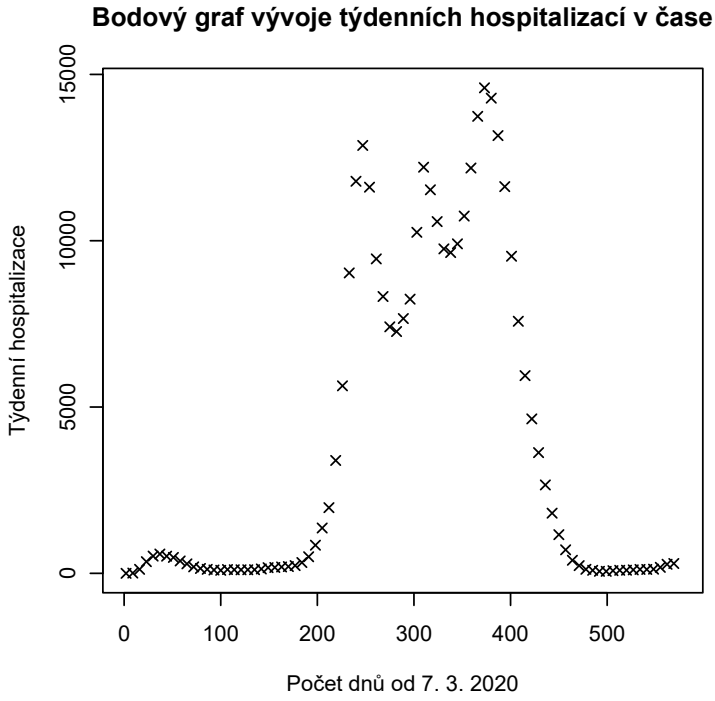
Obrázek 11: Bodový graf pacientů na icu



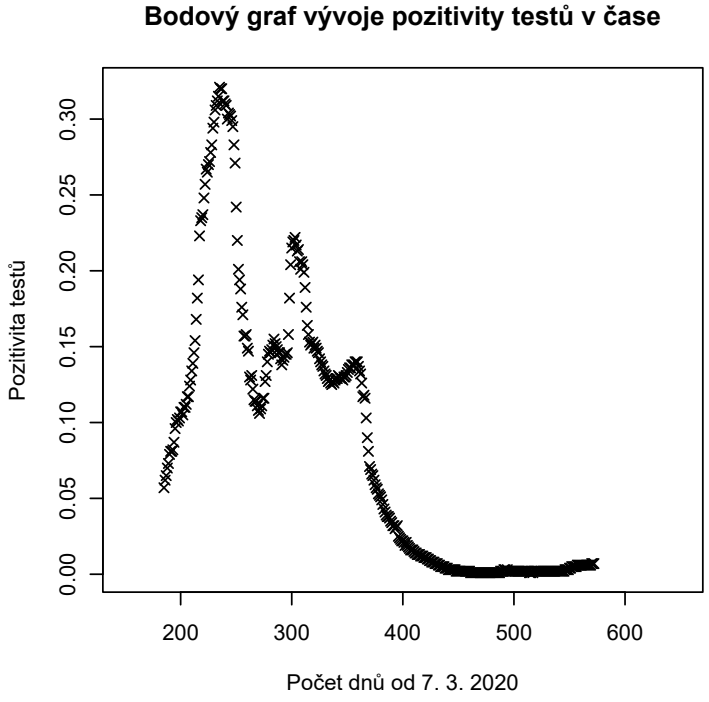
Obrázek 12: Bodový graf hospitalizovaných pacientů



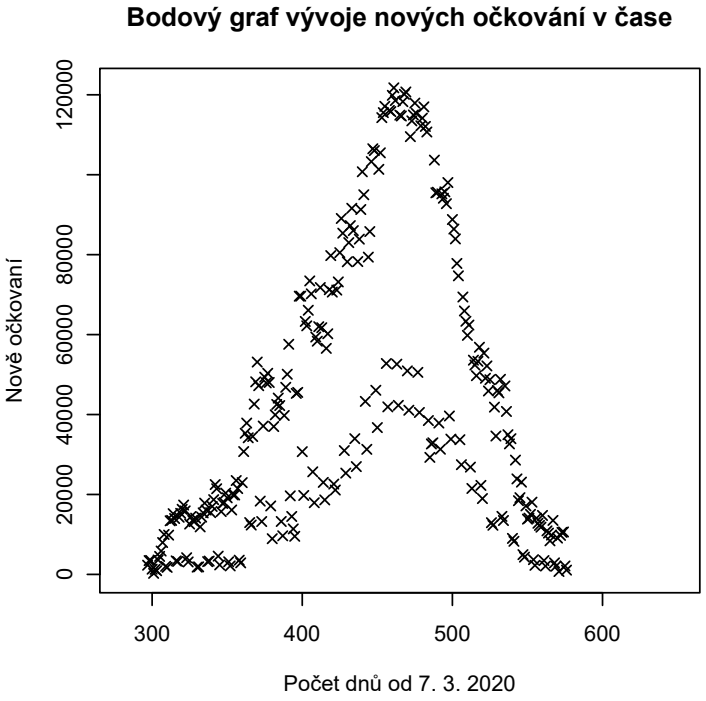
Obrázek 13: Bodový graf týdenních přírůstků na icu



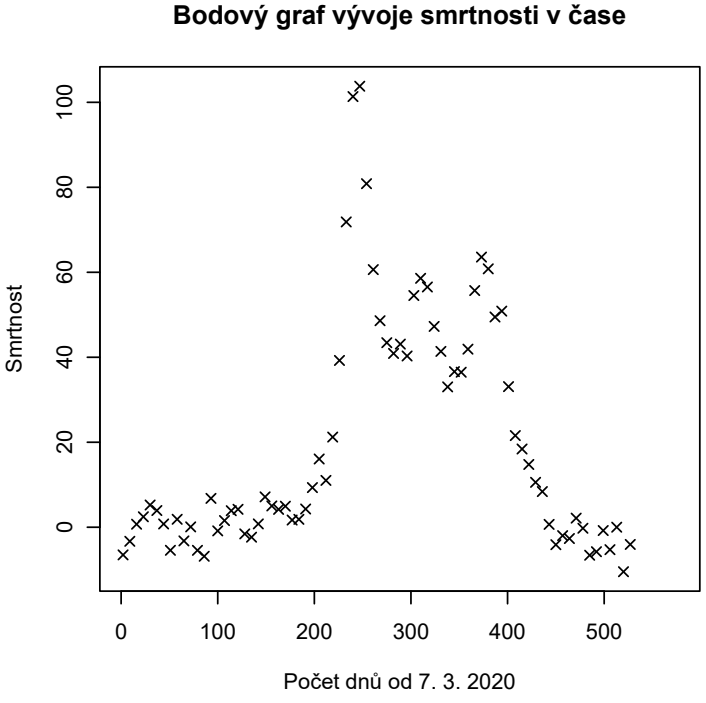
Obrázek 14: Bodový graf týdenních hospitalizací



Obrázek 15: Bodový graf pozitivity testů

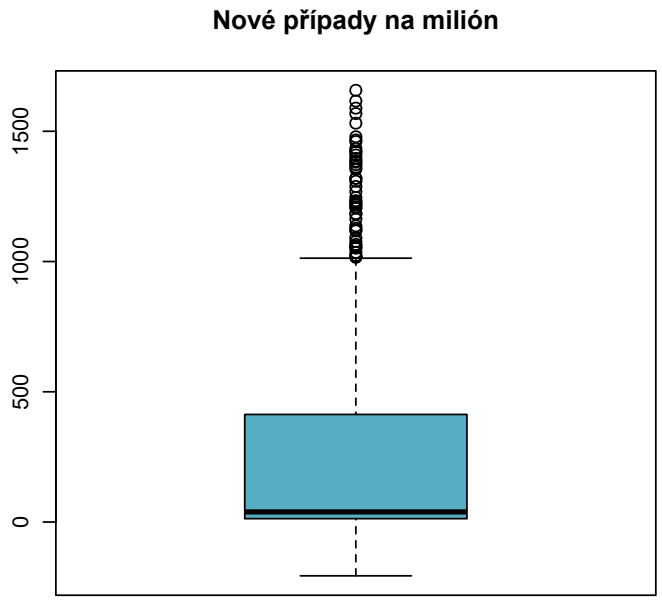


Obrázek 16: Bodový graf nových očkování

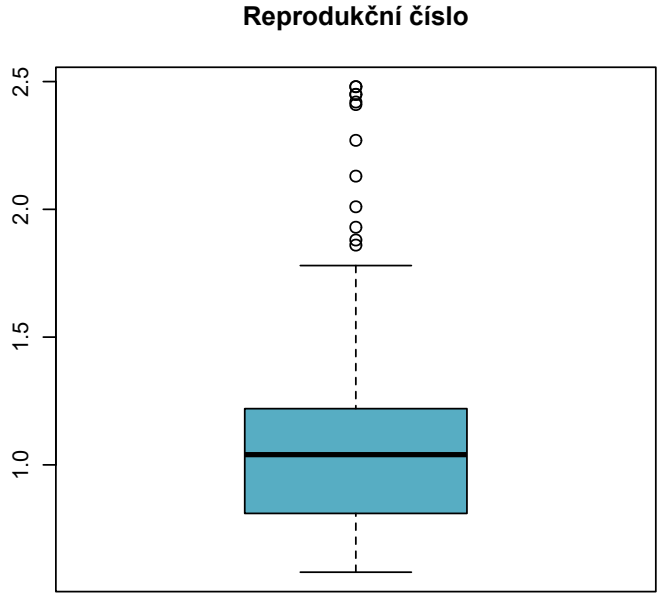


Obrázek 17: Bodový graf smrtnosti

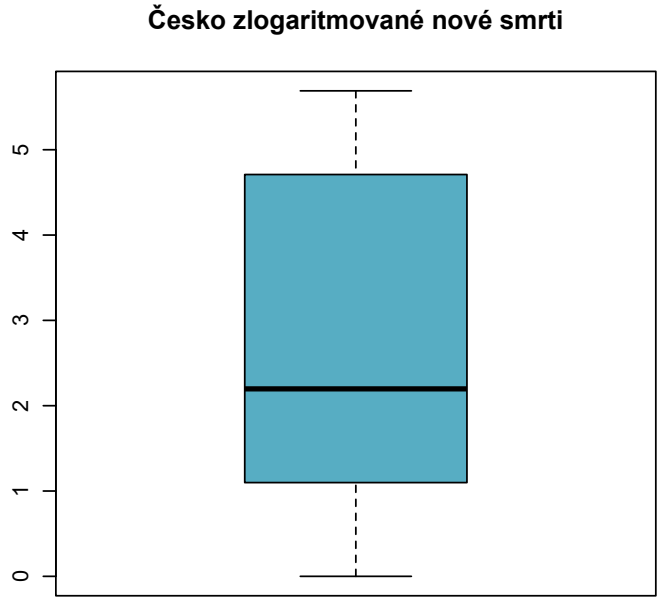
3.3 Boxplot



Obrázek 18: Boxplot graf pro nové případy na milión

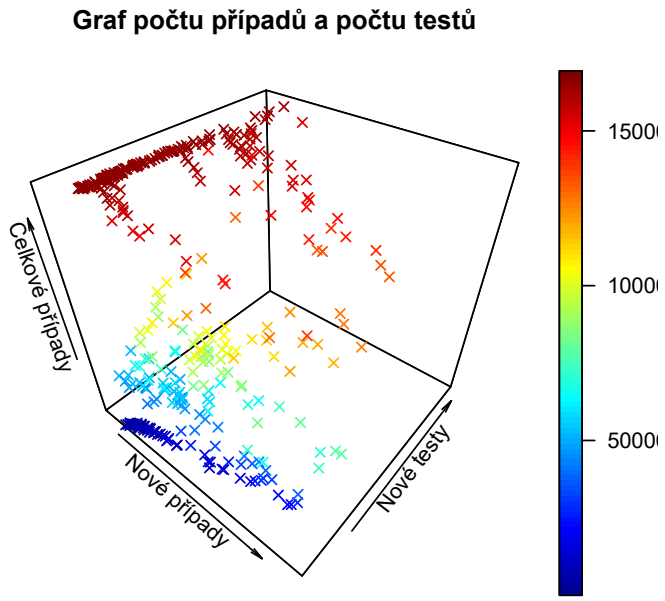


Obrázek 19: Boxplot graf pro reprodukční číslo

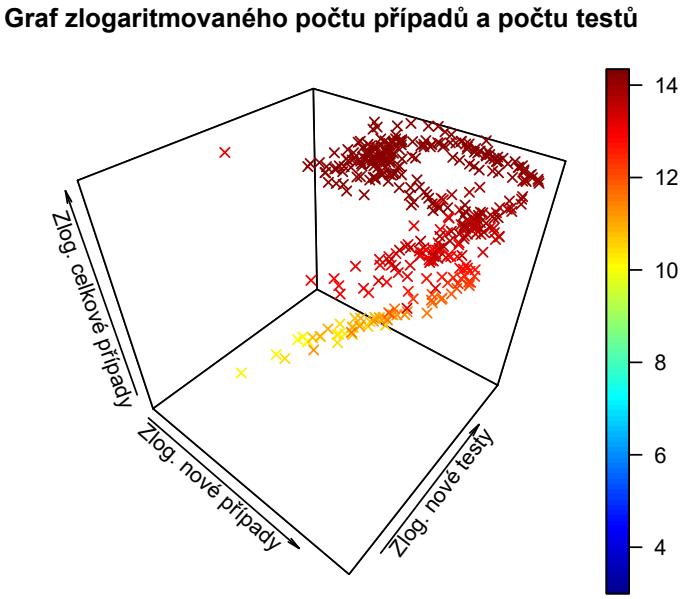


Obrázek 20: Boxplot graf pro zlogaritmované nové smrti

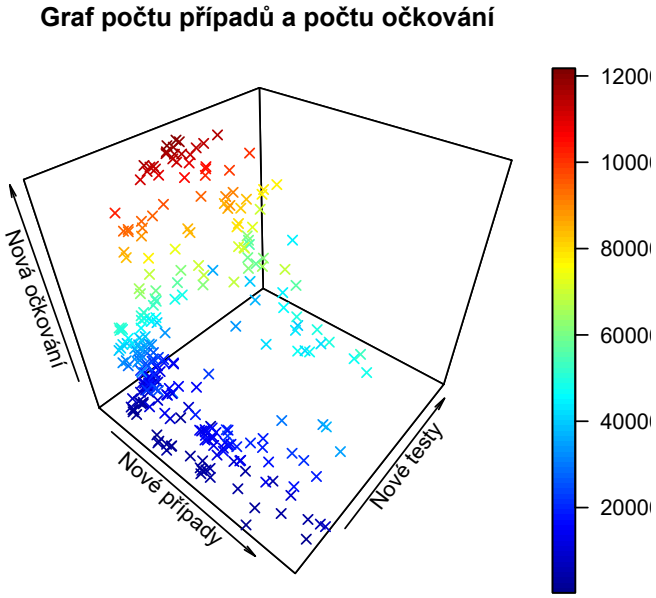
3.4 3D graf



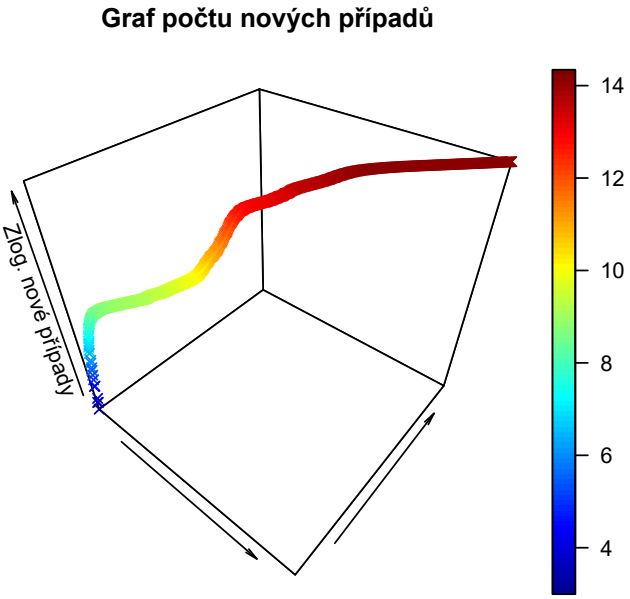
Obrázek 21: 3D graf počtu případů a počtu testů



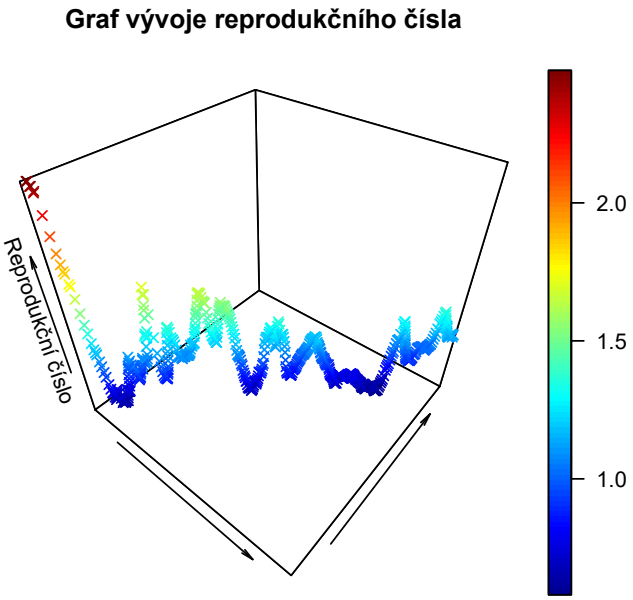
Obrázek 22: 3D graf zlogaritmovaných počtu případů a počtu testů



Obrázek 23: 3D graf počtu případů a počtu nových očkování

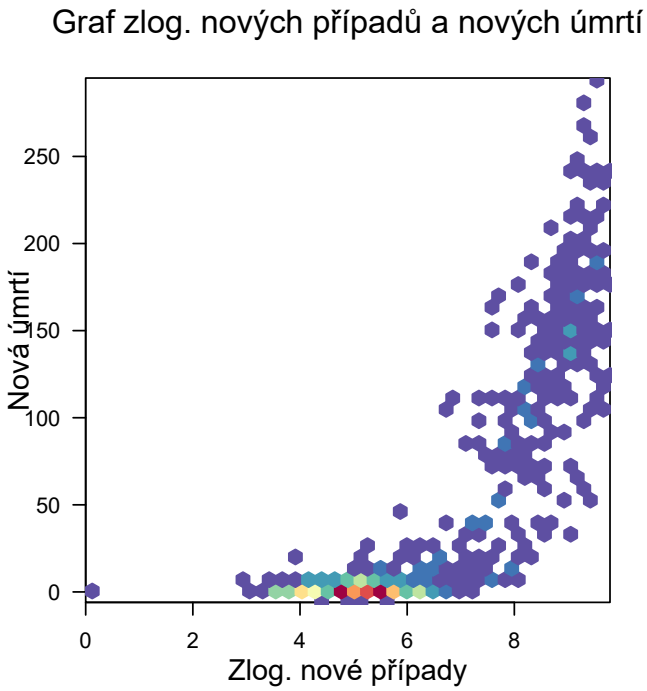


Obrázek 24: 3D graf počtu nových případů



Obrázek 25: 3D graf reprodukčního čísla

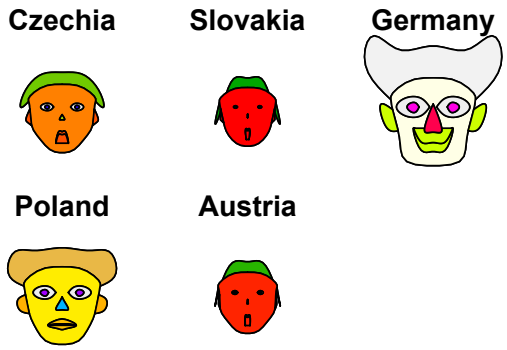
3.5 Hexbin



Obrázek 26: Hexbin graf nových zlog. nových případů a nových úmrtí

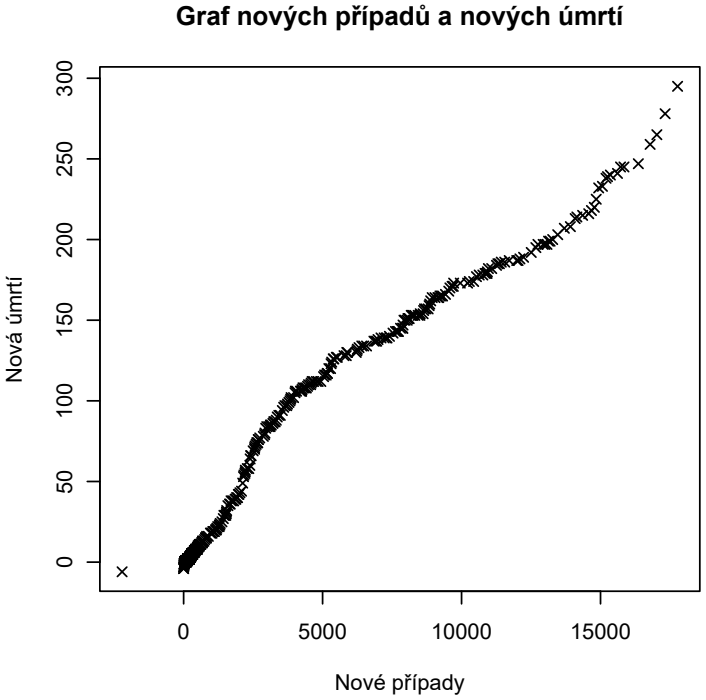
3.6 Chernoff faces

```
effect of variables:
 modified item       Var
 "height of face   " "countriesMeanNewCases"
 "width of face    " "countriesMeanTotalCases"
 "structure of face" "countriesPopulation"
 "height of mouth  " "countriesMeanNewCases"
 "width of mouth   " "countriesMeanTotalCases"
 "smiling          " "countriesPopulation"
 "height of eyes   " "countriesMeanNewCases"
 "width of eyes    " "countriesMeanTotalCases"
 "height of hair   " "countriesPopulation"
 "width of hair    "  "countriesMeanNewCases"
 "style of hair    "  "countriesMeanTotalCases"
 "height of nose   "  "countriesPopulation"
 "width of nose    "  "countriesMeanNewCases"
 "width of ear     "  "countriesMeanTotalCases"
 "height of ear    "  "countriesPopulation"
```

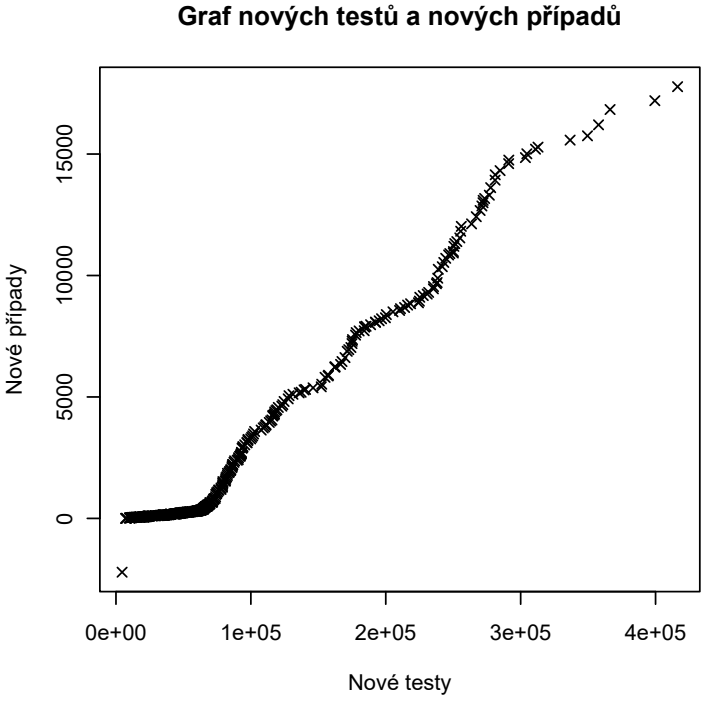


Obrázek 27: Chernoff faces graf tabulky popisné statistiky

3.7 QQPlot



Obrázek 28: QQPlot graf nových případů a nových úmrtí



Obrázek 29: QQPlot graf nových testů a nových případů

4 TESTOVÁNÍ STATISTICKÝCH HYPOTÉZ

4.1 Jednovýběrový Studentův test vůči střední hodnotě

```
        One Sample t-test

data:  data_czech$new_cases
t = -2.0168, df = 575, p-value = 0.04418
alternative hypothesis: true mean is not equal to 3300
95 percent confidence interval:
 2590.550 3290.603
sample estimates:
mean of x
 2940.576

        One Sample t-test

data:  data_czech$new_cases_smoothed
t = 18.184, df = 575, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 2619.672 3254.109
sample estimates:
mean of x
  2936.89

        One Sample t-test

data:  data_czech$new_cases_per_million
t = 16.5, df = 575, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 241.5531 306.8289
sample estimates:
mean of x
  274.191
```

```
        One Sample t-test

data:  data_czech$new_cases_smoothed_per_million
t = 18.184, df = 575, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 244.2687 303.4260
sample estimates:
mean of x
 273.8474

        One Sample t-test

data:  data_czech$hosp_patients
t = 19.019, df = 568, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 2125.690 2615.294
sample estimates:
mean of x
 2370.492

        One Sample t-test

data:  data_czech$hosp_patients_per_million
t = 19.019, df = 568, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 198.2078 243.8604
sample estimates:
mean of x
 221.0341
```

4.2 Dvouvýběrový Studentův test

```
        Welch Two Sample t-test

data:  p1 and p2
t = -4.518, df = 537.03, p-value = 7.683e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2272.4359  -895.1752
sample estimates:
mean of x mean of y
 2148.674  3732.479

        Welch Two Sample t-test

data:  data_czech$new_cases_per_million and data_slovakia$new_cases_per_million
t = 7.7283, df = 817.84, p-value = 3.194e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 105.8863 177.9857
sample estimates:
mean of x mean of y
  274.191   132.255

        Two Sample t-test

data:  data_czech$new_cases_per_million and data_germany$new_cases_per_million
t = 10.844, df = 1150, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 152.3817 219.7075
sample estimates:
mean of x mean of y
274.19103  88.14644
```

```
        Two Sample t-test

data:  data_czech$new_cases_per_million and data_poland$new_cases_per_million
t = 7.5404, df = 1150, p-value = 9.477e-14
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 103.9326 177.0431
sample estimates:
mean of x mean of y
 274.1910  133.7032

        Two Sample t-test

data:  data_czech$new_cases_per_million and data_austria$new_cases_per_million
t = 7.2242, df = 1150, p-value = 9.157e-13
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  95.01377 165.86690
sample estimates:
mean of x mean of y
 274.1910  143.7507
```

4.3 Wilcox test

```
        Wilcoxon rank sum test with continuity correction

data:  data_czech$new_cases_per_million and data_slovakia$new_cases_per_million
W = 205293, p-value = 2.97e-12
alternative hypothesis: true location shift is not equal to 0

        Wilcoxon rank sum test with continuity correction

data:  data_czech$new_cases_per_million and data_germany$new_cases_per_million
W = 188720, p-value = 5.261e-05
alternative hypothesis: true location shift is not equal to 0
```

4.4 Fisherův test

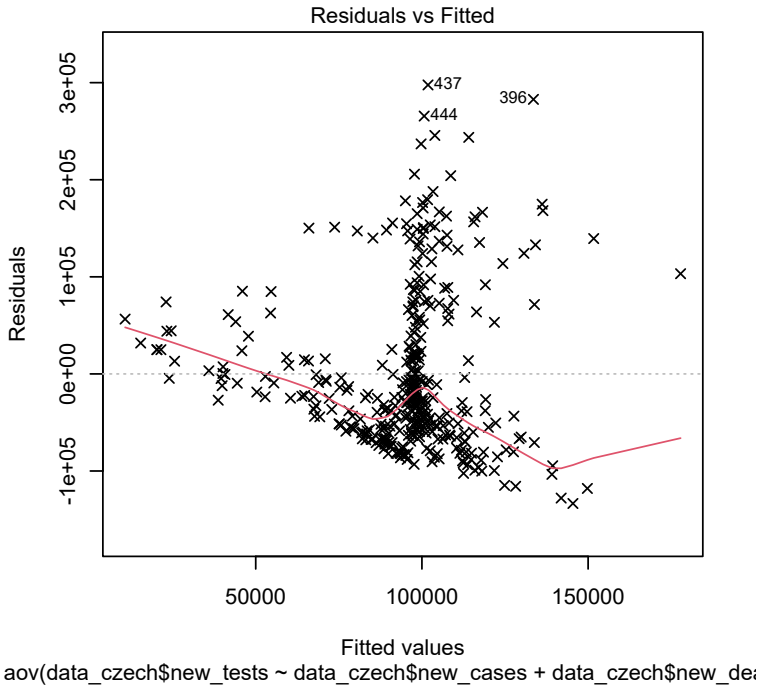
```
        F test to compare two variances

data:  lm(data_czech$new_cases_per_million ~ 1) and lm(data_slovakia$new_cases_per_million ~ 1)
F = 4.5141, num df = 575, denom df = 575, p-value < 2.2e-16
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 3.832723 5.316656
sample estimates:
ratio of variances
          4.514119
```

4.5 Shapiro Wilk test

```
        Shapiro-Wilk normality test

data:  data_czech$new_cases
W = 0.72003, p-value < 2.2e-16

        Shapiro-Wilk normality test

data:  data_czech$new_tests
W = 0.82915, p-value < 2.2e-16
```

5 ANOVA

```
                       Df    Sum Sq   Mean Sq F value   Pr(>F)
data_czech$new_cases    1 6.973e+10 6.973e+10   10.81   0.0011 **
data_czech$new_deaths   1 1.218e+11 1.218e+11   18.88 1.78e-05 ***
Residuals             391 2.522e+12 6.451e+09
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
182 observations deleted due to missingness
```



Obrázek 30: Anova graf nových testů, případů a úmrtí

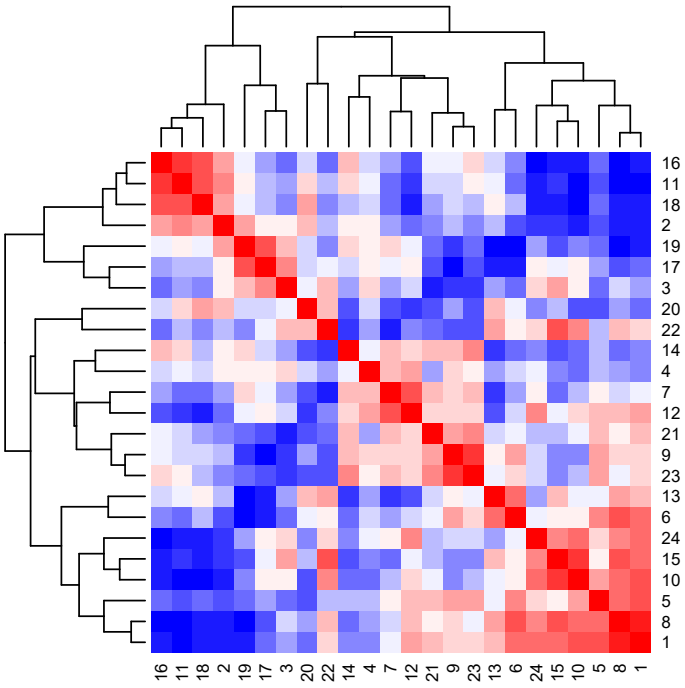
6 VARIANCE

```
     Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
-61749419 -61749419 -61749419 -61749419 -61749419 -61749419
```

7 KORELACE

7.1 Korelační matice

```
             [,1]        [,2]         [,3]         [,4]        [,5]        [,6]
 [1,]  1.00000000 -0.62274409 -0.129186615 -0.285154793  0.66579693  0.57478261
 [2,] -0.62274409  1.00000000  0.151838159  0.209449286 -0.43366681 -0.44705372
 [3,] -0.12918662  0.15183816  1.000000000  0.250435635 -0.36067001 -0.33492826
 [4,] -0.28515479  0.20944929  0.250435635  1.000000000 -0.10668415 -0.03613412
 [5,]  0.66579693 -0.43366681 -0.360670011 -0.106684148  1.00000000  0.50445750
 [6,]  0.57478261 -0.44705372 -0.334928261 -0.036134119  0.50445750  1.00000000
 [7,]  0.05132667 -0.14357190 -0.163185379  0.386541958  0.15705896 -0.13049153
 [8,]  0.89739130 -0.55838227  0.004784689 -0.171528226  0.57403785  0.69391304
 [9,]  0.30782609 -0.09088933 -0.482383691  0.227688483  0.39791260  0.43652174
[10,]  0.72869565 -0.63448577  0.187472832 -0.359599784  0.45966516  0.17739130
[11,] -0.66869565  0.54794522 -0.131361474  0.079233851 -0.46488368 -0.32956522
[12,]  0.42782609 -0.34007394  0.016093955  0.404005569  0.37442923  0.02869565
[13,]  0.38695652 -0.12002610 -0.172683792 -0.201567434  0.10784954  0.64347826
[14,] -0.27478261  0.13785606 -0.133971305  0.135394108 -0.09306371 -0.34956522
[15,]  0.61043478 -0.47358122  0.410613349 -0.280365934  0.17612525  0.16434783
[16,] -0.59304348  0.45749077 -0.321879108  0.026121050 -0.31919983 -0.23913043
[17,] -0.31652174  0.21743858  0.533710359  0.180235243 -0.17090672 -0.62434783
[18,] -0.59478261  0.47271147 -0.263157920  0.048323942 -0.32398348 -0.08000000
[19,] -0.60608696  0.46879758  0.355371934  0.154549544 -0.33529029 -0.68608696
[20,] -0.34398783  0.31165724  0.099412663 -0.006749405 -0.46889952  0.13089802
[21,]  0.32347826 -0.23265928 -0.586341943 -0.168916121  0.30876278  0.09043478
[22,]  0.23483367 -0.11048282  0.382423329 -0.190289684 -0.11439756  0.20656665
[23,]  0.25701240 -0.24880383 -0.511638037  0.139995729  0.47498913  0.25179387
[24,]  0.59143293 -0.52522836  0.291929526  0.050729401  0.30230535  0.09088933
             [,7]         [,8]          [,9]       [,10]        [,11]
 [1,]  0.05132667  0.897391304  0.3078260870  0.72869565 -0.668695652
 [2,] -0.14357190 -0.558382270 -0.0908893259 -0.63448577  0.547945218
 [3,] -0.16318538  0.004784689 -0.4823836907  0.18747283 -0.131361474
 [4,]  0.38654196 -0.171528226  0.2276884833 -0.35959978  0.079233851
 [5,]  0.15705896  0.574037848  0.3979125991  0.45966516 -0.464883681
 [6,] -0.13049153  0.693913043  0.4365217391  0.17739130 -0.329565217
 [7,]  1.00000000 -0.029578080  0.2801218186 -0.05480644 -0.334058318
 [8,] -0.02957808  1.000000000  0.3043478261  0.62956522 -0.664347826
 [9,]  0.28012182  0.304347826  1.0000000000 -0.23304348  0.008695652
[10,] -0.05480644  0.629565217 -0.2330434783  1.00000000 -0.646086957
[11,] -0.33405832 -0.664347826  0.0086956522 -0.64608696  1.000000000
[12,]  0.68595048  0.310434783  0.2269565217  0.28782609 -0.530434783
```

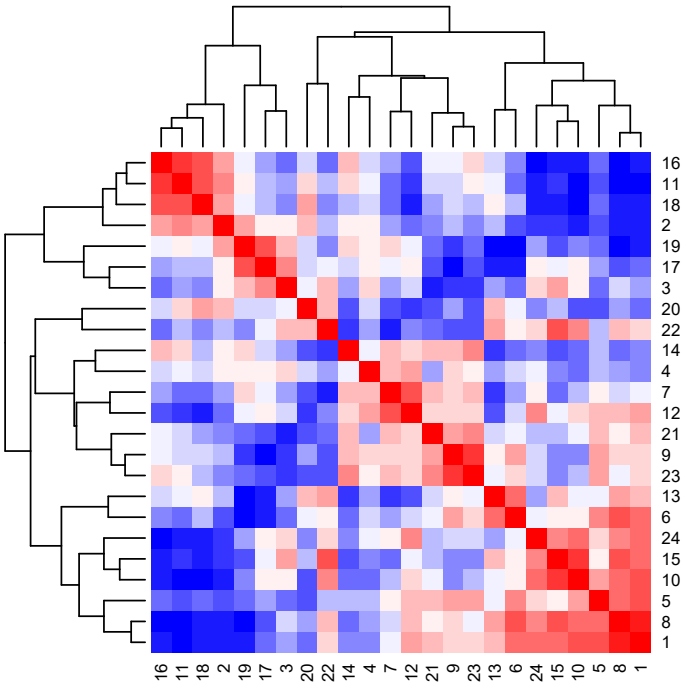


Obrázek 31: Heatmap graf korelační matice

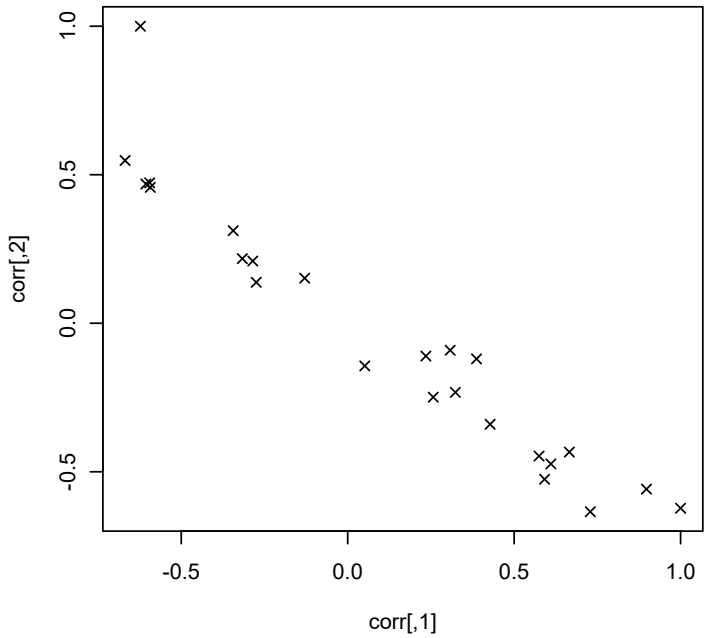
8 KOVARIANCE

8.1 Kovarianční matice

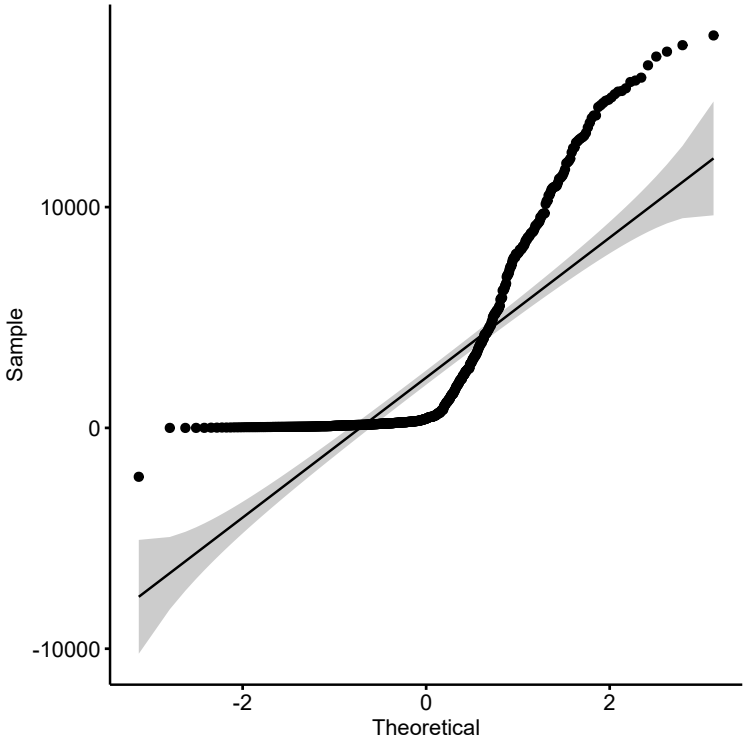
```
      V1                V2                V3                V4
 Min.   :-33.435   Min.   :-31.717   Min.   :-29.3043   Min.   :-17.957
 1st Qu.:-16.168   1st Qu.:-21.842   1st Qu.:-13.8859   1st Qu.: -8.467
 Median : 12.293   Median : -5.761   Median : -3.1087   Median :  2.473
 Mean   :  5.465   Mean   : -2.097   Mean   :  0.1916   Mean   :  3.088
 3rd Qu.: 28.946   3rd Qu.: 12.046   3rd Qu.: 13.0217   3rd Qu.:  9.364
 Max.   : 50.000   Max.   : 49.978   Max.   : 49.9565   Max.   : 49.870
      V5                V6                V7                V8
 Min.   :-23.435   Min.   :-34.304   Min.   :-29.1739   Min.   :-33.217
 1st Qu.:-16.016   1st Qu.:-13.087   1st Qu.:-10.8207   1st Qu.:-17.967
 Median :  6.620   Median :  4.533   Median :  0.5435   Median :  8.315
 Mean   :  4.815   Mean   :  3.619   Mean   :  2.1472   Mean   :  5.156
 3rd Qu.: 20.663   3rd Qu.: 14.897   3rd Qu.: 14.7663   3rd Qu.: 25.891
 Max.   : 49.978   Max.   : 50.000   Max.   : 49.9565   Max.   : 50.000
      V9               V10               V11               V12
 Min.   :-34.435   Min.   :-36.435   Min.   :-33.435   Min.   :-31.174
 1st Qu.:-10.658   1st Qu.:-17.337   1st Qu.:-23.321   1st Qu.:-14.668
 Median :  8.826   Median :  6.196   Median : -1.522   Median : 10.457
 Mean   :  4.632   Mean   :  3.081   Mean   : -1.861   Mean   :  4.583
 3rd Qu.: 16.087   3rd Qu.: 23.522   3rd Qu.:  9.120   3rd Qu.: 16.321
 Max.   : 50.000   Max.   : 50.000   Max.   : 50.000   Max.   : 50.000
     V13               V14               V15               V16
 Min.   :-35.739   Min.   :-25.9348   Min.   :-30.8696   Min.   :-33.783
 1st Qu.: -8.989   1st Qu.:-17.0652   1st Qu.:-17.2446   1st Qu.:-17.913
 Median :  3.685   Median : -3.0435   Median :  0.4348   Median : -5.250
 Mean   :  2.385   Mean   :  0.5525   Mean   :  2.7237   Mean   : -2.403
 3rd Qu.: 17.924   3rd Qu.: 14.1957   3rd Qu.: 21.7337   3rd Qu.:  8.435
 Max.   : 50.000   Max.   : 50.0000   Max.   : 50.0000   Max.   : 50.000
     V17               V18               V19               V20
 Min.   :-34.4348   Min.   :-36.435   Min.   :-35.739   Min.   :-27.087
 1st Qu.:-17.0543   1st Qu.:-19.696   1st Qu.:-19.120   1st Qu.:-19.397
 Median : -0.6956   Median : -5.652   Median : -3.815   Median : -2.185
 Mean   : -1.3895   Mean   : -3.584   Mean   : -3.570   Mean   : -2.415
 3rd Qu.:  9.1413   3rd Qu.:  6.598   3rd Qu.:  8.679   3rd Qu.:  8.489
 Max.   : 50.0000   Max.   : 50.000   Max.   : 50.000   Max.   : 49.978
     V21               V22               V23               V24
 Min.   :-29.304   Min.   :-29.174   Min.   :-25.565   Min.   :-33.783
 1st Qu.: -9.234   1st Qu.:-14.495   1st Qu.:-12.223   1st Qu.: -9.245
```



Obrázek 32: Heatmap graf korelační matice



Obrázek 33: Graf korelační matice



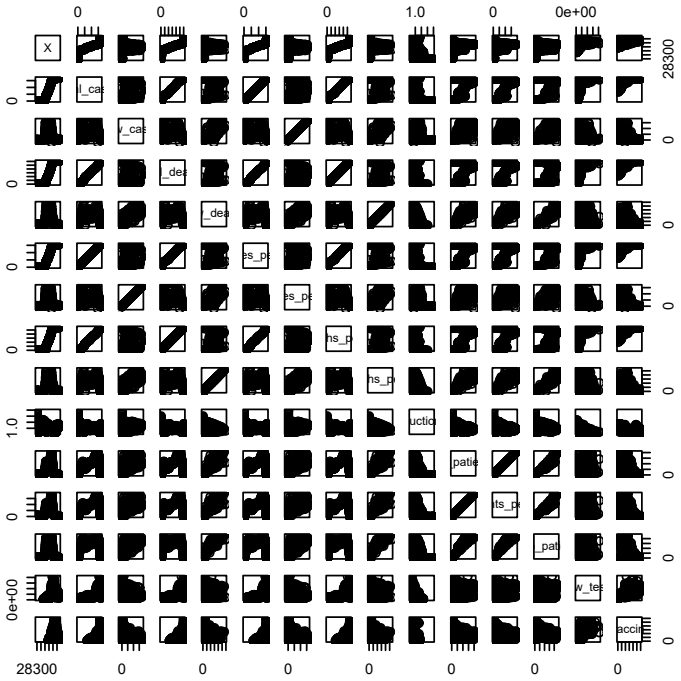
Obrázek 34: GGQQPlot graf korelační matice

9 TESTOVÁNÍ V KONTINGENČNÍCH TABULKÁCH

9.1 Pearsonův Chí-kvadrát test

```
        Pearson's Chi-squared test

data:  data_czech$new_tests and data_czech$new_cases
X-squared = 147356, df = 146982, p-value = 0.245
```

10 PAIRS



Obrázek 35: Grafy párů

ZÁVĚR

POUŽITÁ LITERATURA

[1] Our World in Data *Data on COVID-19 (coronavirus)* [online]. 2021 [cit. 2021-11-18]. Dostupné z: `https://github.com/owid/covid-19-data/tree/master/public/data`

SEZNAM PŘÍLOH

PŘÍLOHA A

Příloha A zahrnuje ZIP soubor, který obsahuje: