COVID-19 outbreak

Eduard Bakstein, eduard.bakstein@nudz.cz, http://bakstein.net, @edbakstein

18.3.2020

Intro

The aim is to do exploratory analyses of the COVID-19 cases in selected countries and see how the models evolve as new cases are reported. Can be used as a quick&dirty basis for your own analyses. Special focus given to Czech Rep., as this is my country of origin.

Source code (R markdown): https://github.com/ebakstein/covid19.git

Resources:

- https://rviews.rstudio.com/2020/03/05/covid-19-epidemiology-with-r/ (basis of this rmd and analysis)
- https://timchurches.github.io/blog/posts/2020-02-18-analysing-covid-19-2019-ncov-outbreak-data-with-r-part-1/

Interesting models and remarks

- https://blog.ephorie.de/epidemiology-how-contagious-is-novel-coronavirus-2019-ncov
- https://arxiv.org/pdf/2002.00418v1.pdf

Data sources:

- https://en.wikipedia.org/w/index.php?title=2020_coronavirus_outbreak_in_the_United_States&oldid=944107102
- https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series

TO DO

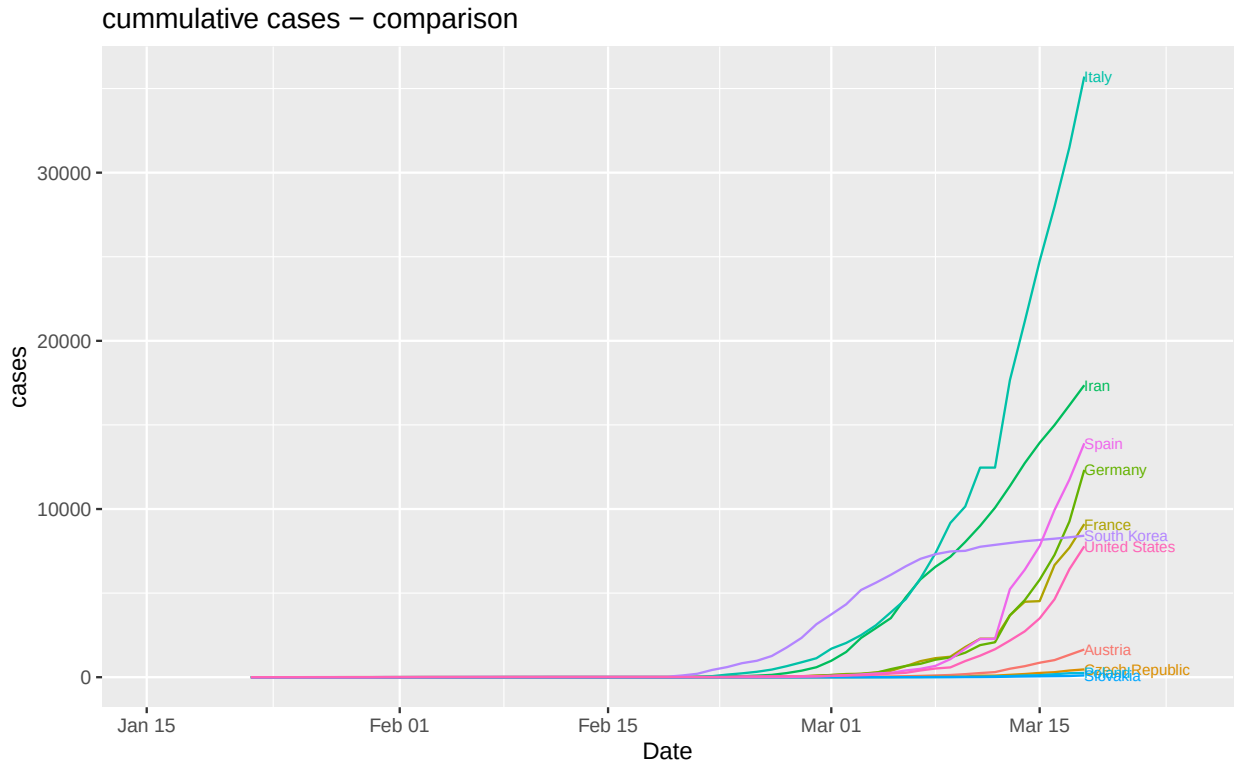
- modelling

Load data

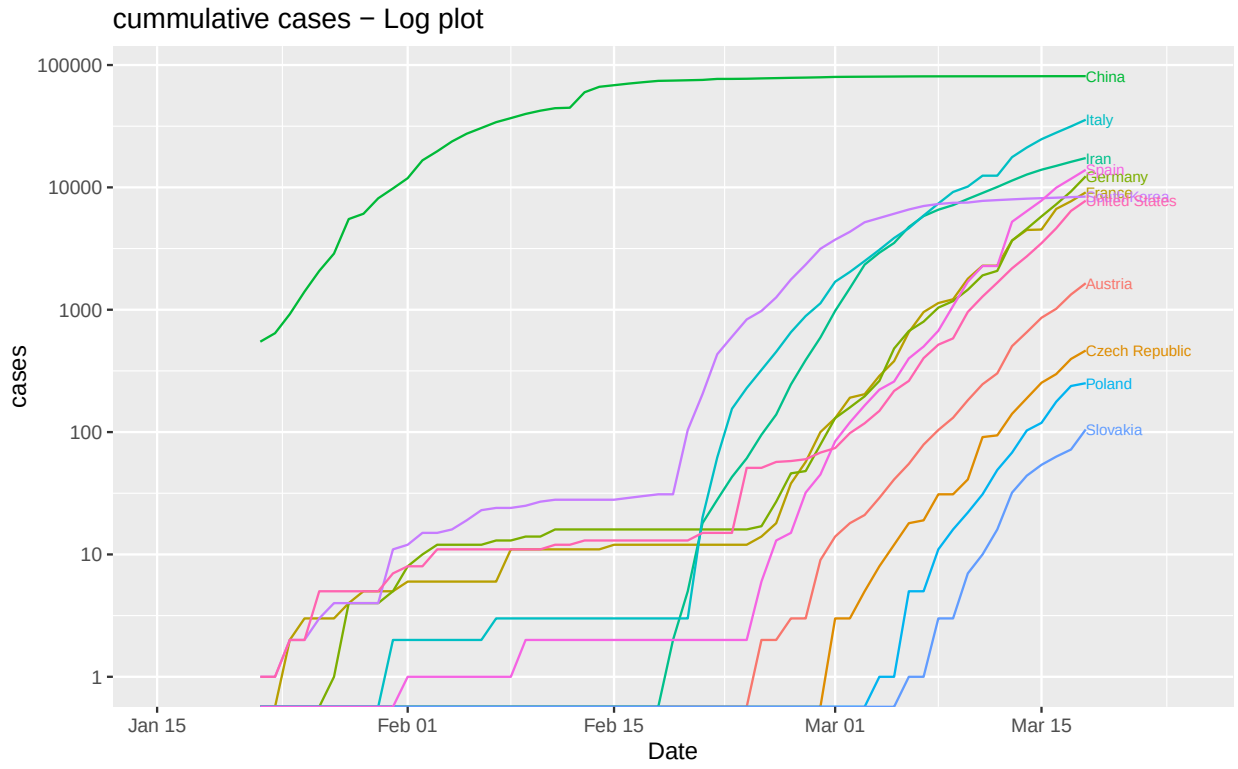
Using the Jons-Hopkins github repository to obtain current data, see https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series (updated daily around 23:59 UTC)

Obtain data about population in each country by scraping the wikipedia page: 'https://en.wikipedia.org/wiki/List_of_countries_by_population_(United_Nations)'

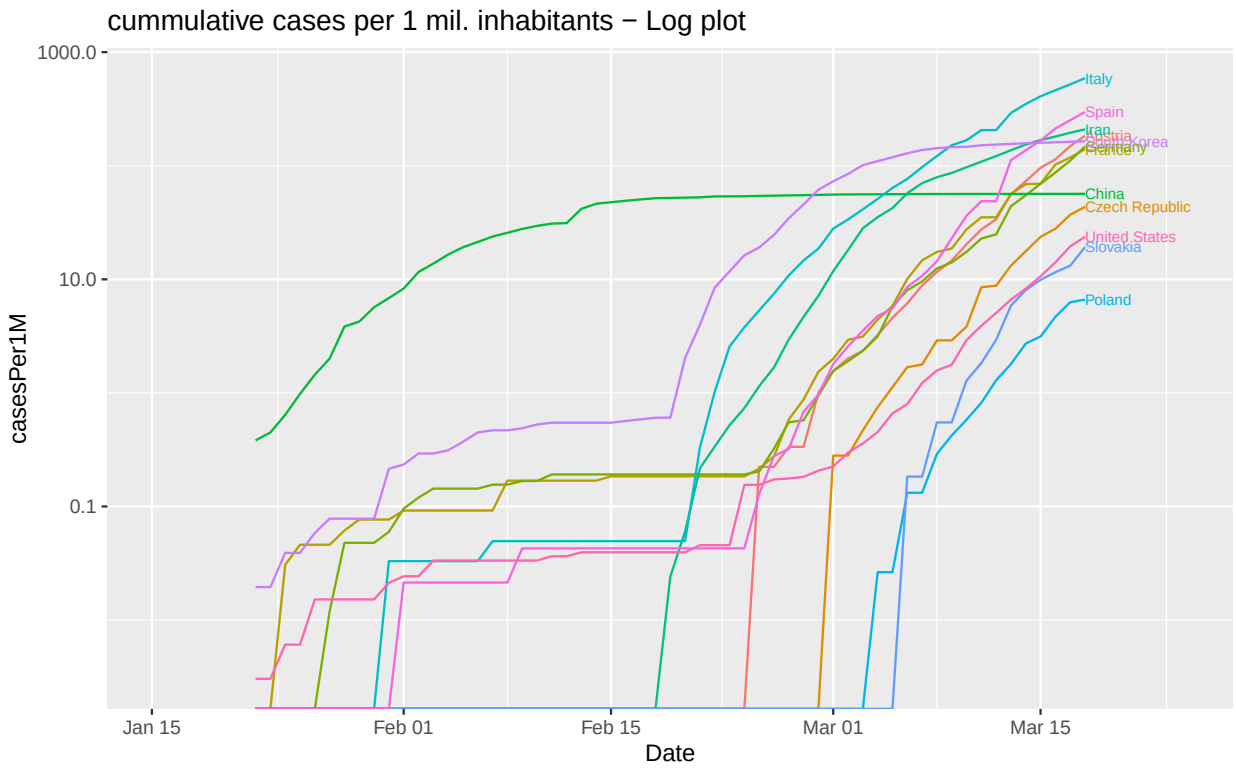
Plot selected countries



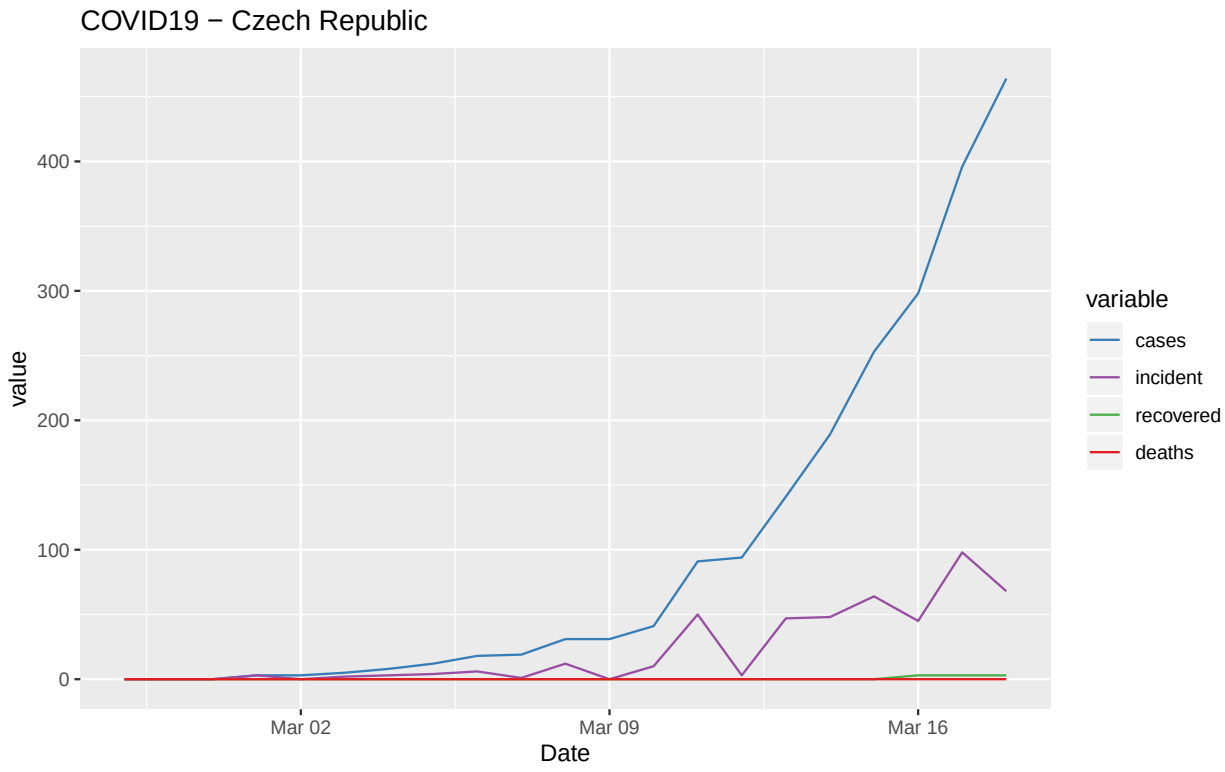
Note: China number of cases too high to be plot on linear y axis with other countries

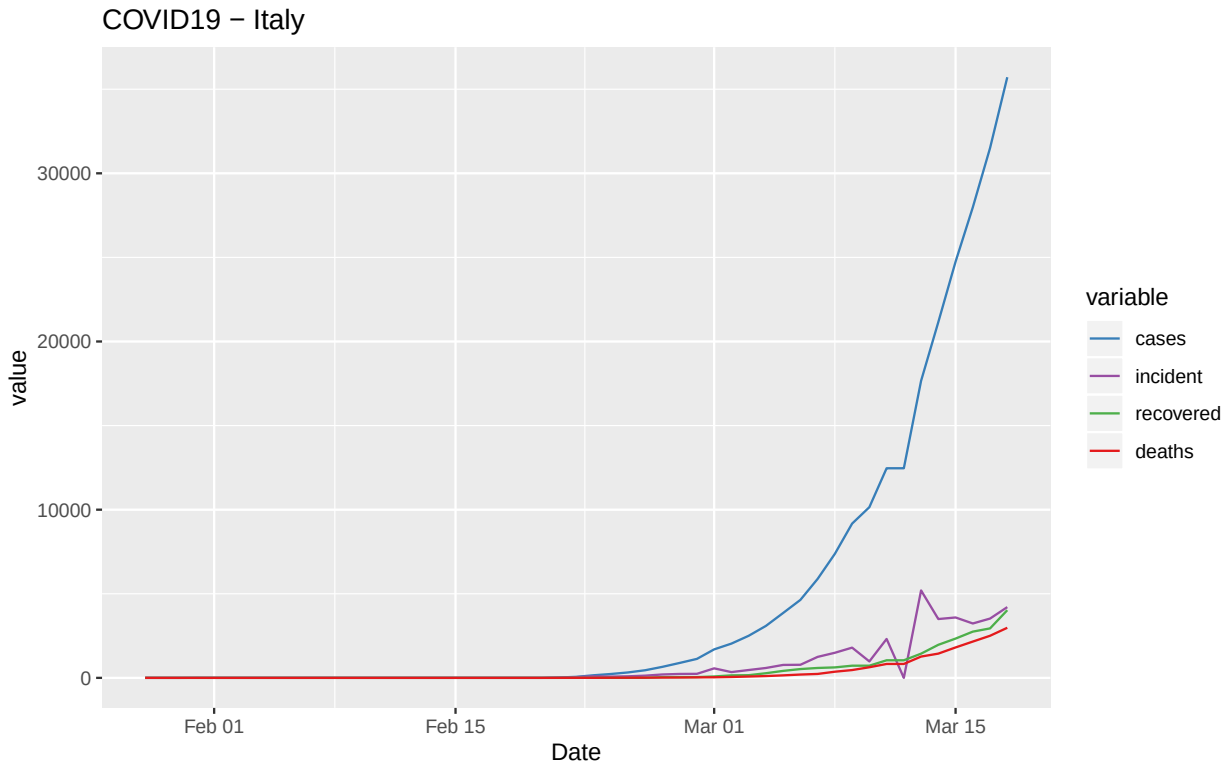


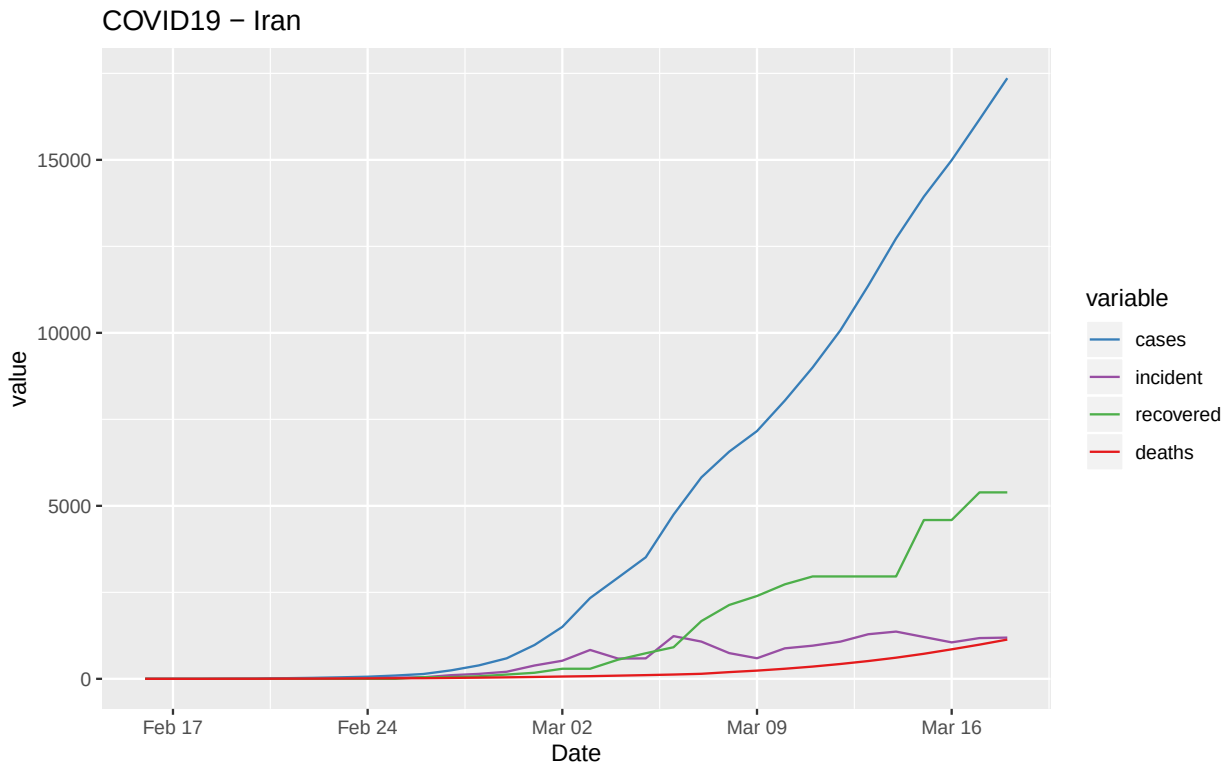
Note: China added to the log plot

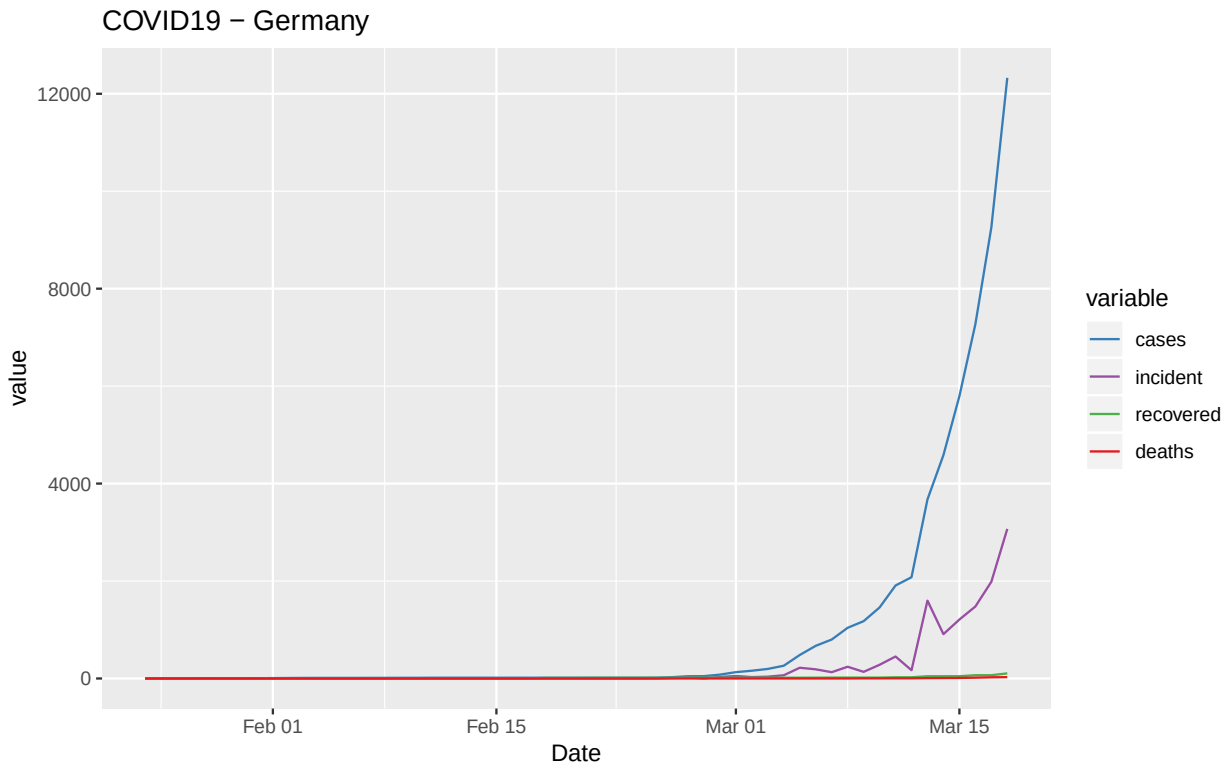

cummulative cases per 1 mil. inhabitants − Log plot
1000.0
10.0
0.1
casesPer1M
Jan 15
Feb 01
Feb 15
Mar 01
Mar 15
Date
Italy
Spain
Iran
Austria
South Korea
France
Germany
China
Czech Republic
United States
Slovakia
Poland

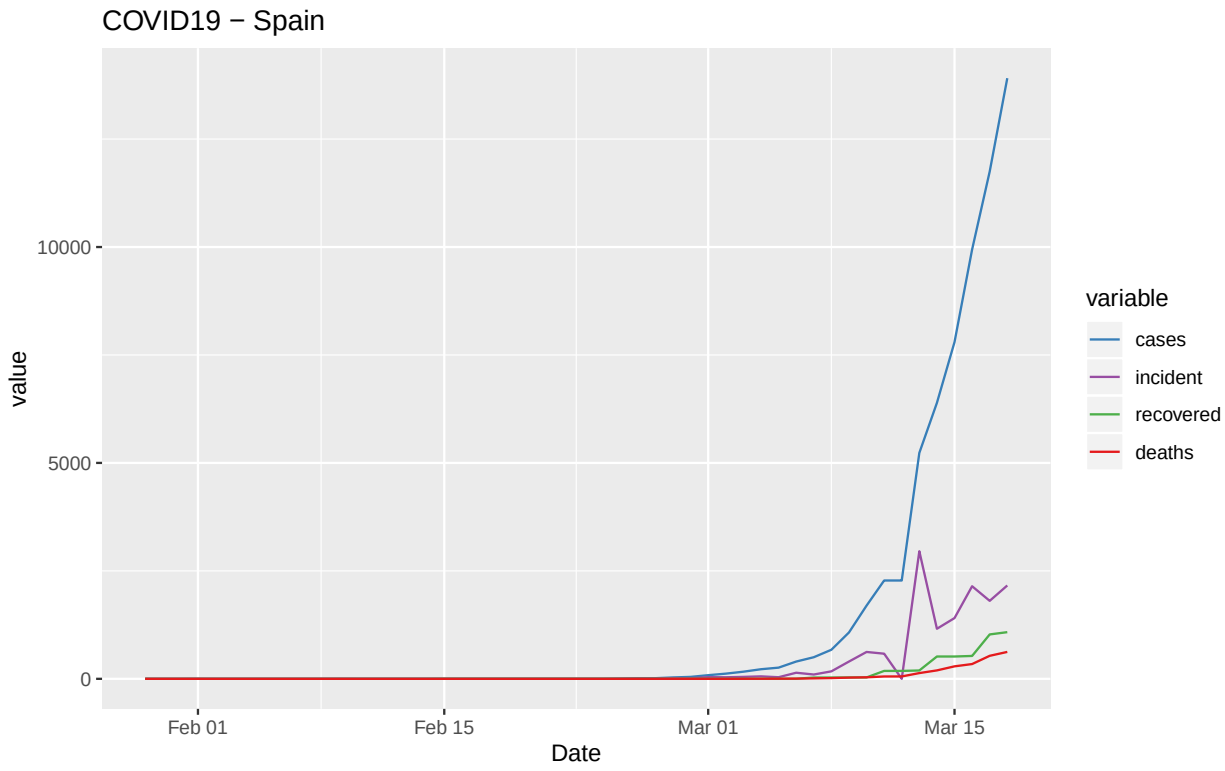
Cummulative stats for Individual selected countries

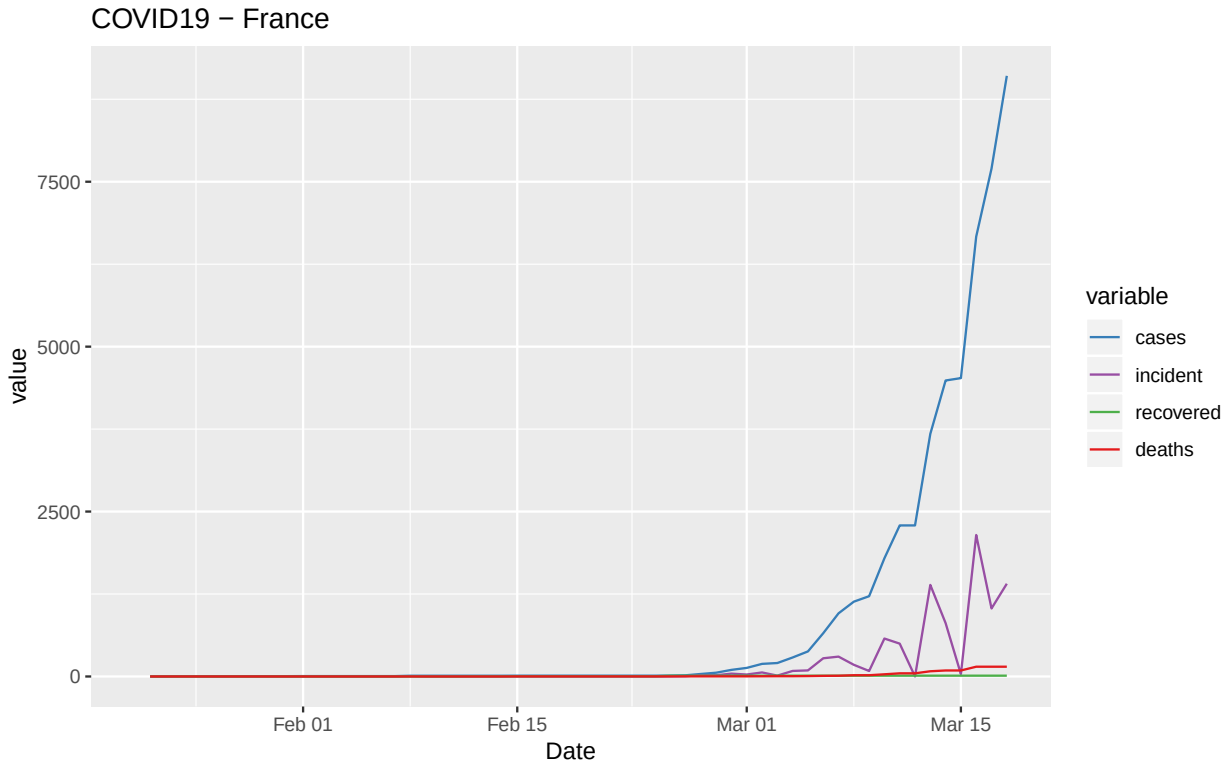


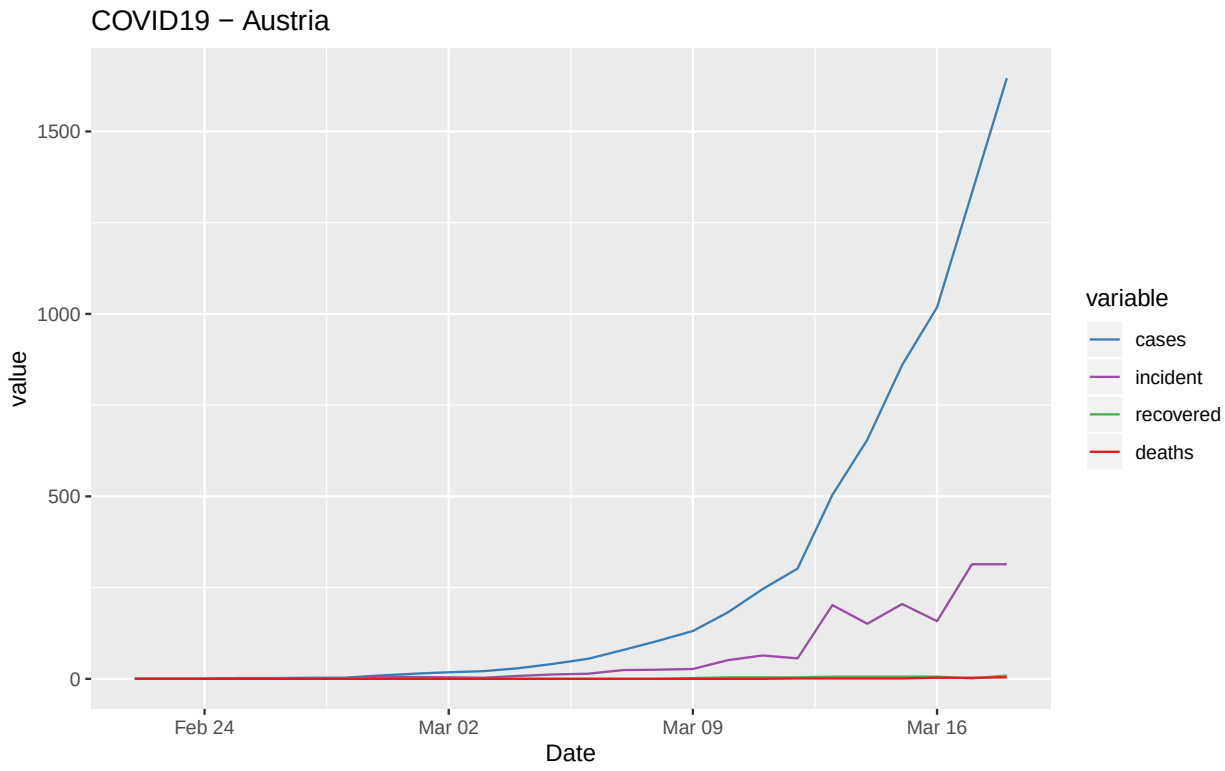


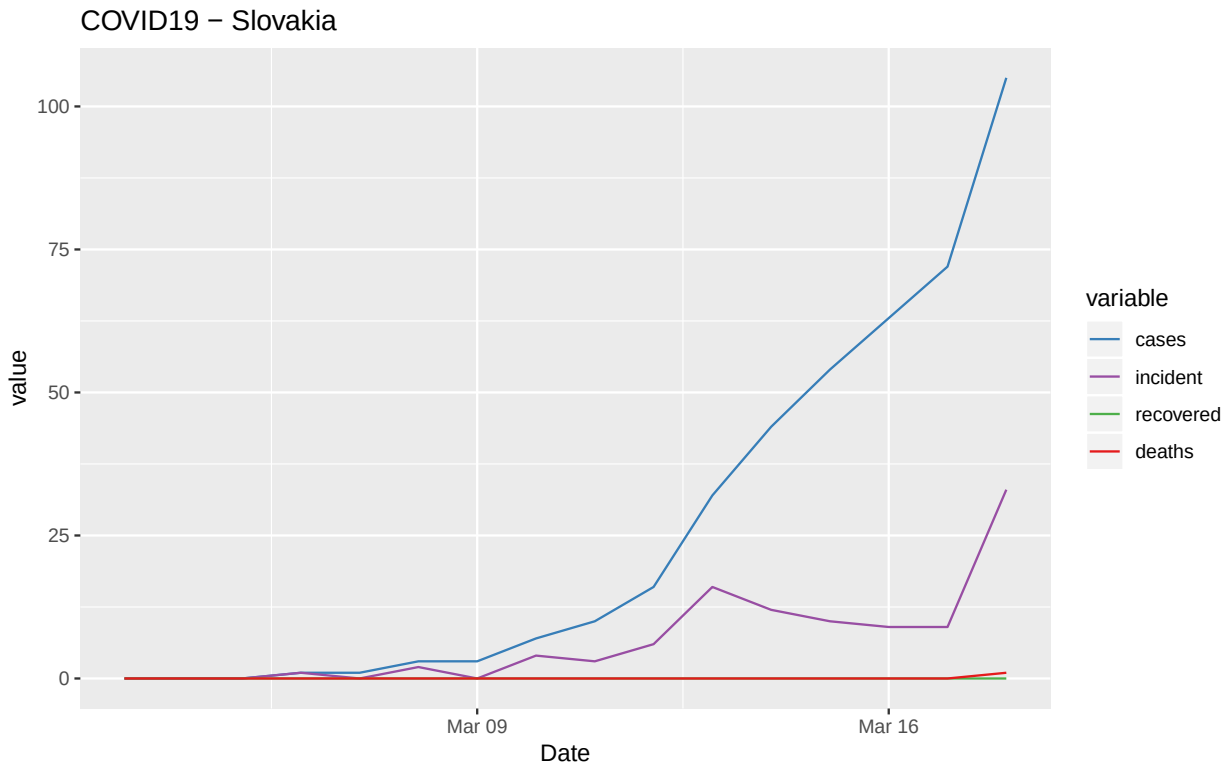

COVID19 − Iran
15000
10000
5000
0
value
Feb 17
Feb 24
Mar 02
Mar 09
Mar 16
Date
variable
cases
incident
recovered
deaths

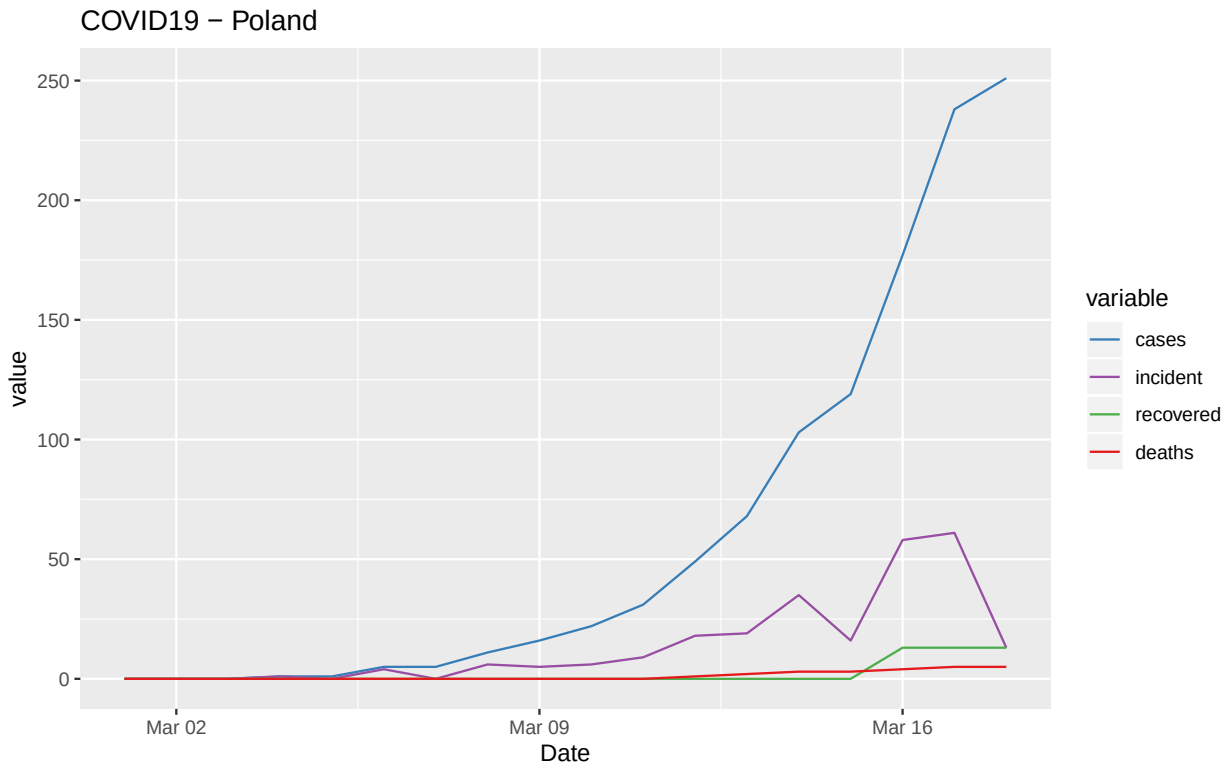

COVID19 − Germany
12000
8000
4000
0
value
Feb 01
Feb 15
Mar 01
Mar 15
Date
variable
cases
incident
recovered
deaths

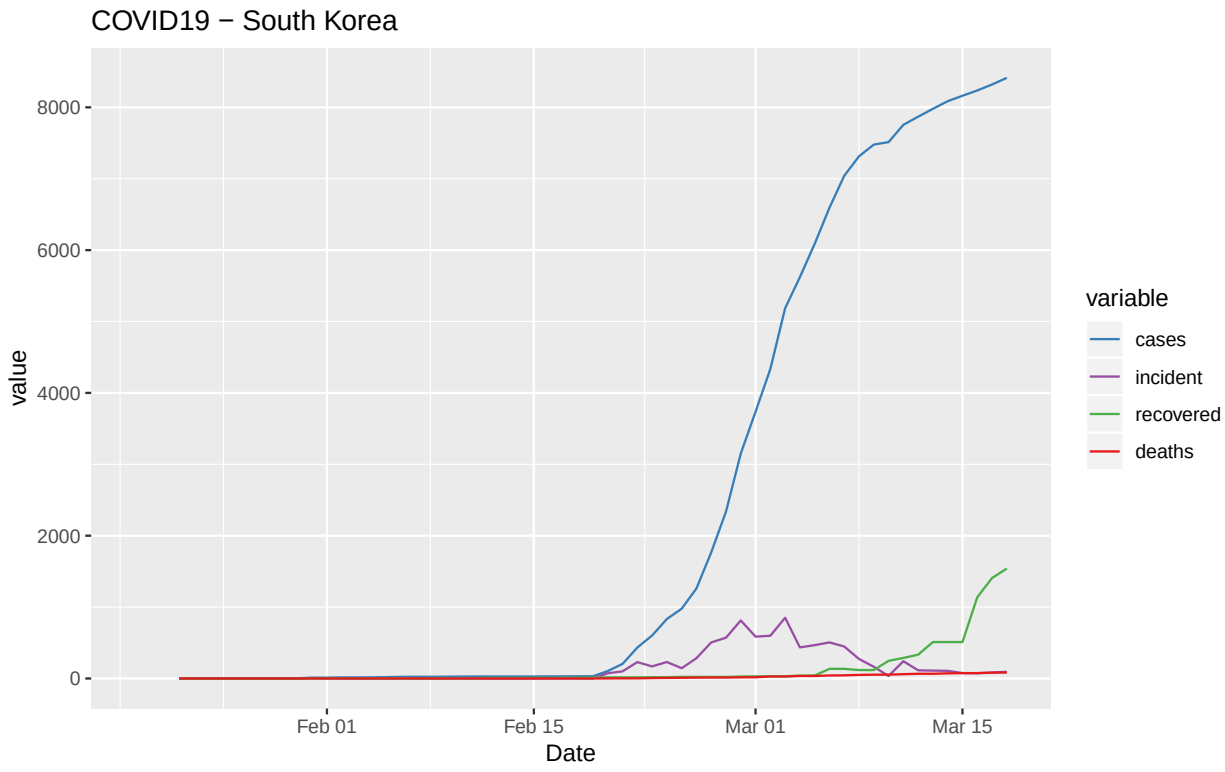

COVID19 − Spain
value
10000
5000
0
Feb 01
Feb 15
Mar 01
Mar 15
Date
variable
cases
incident
recovered
deaths

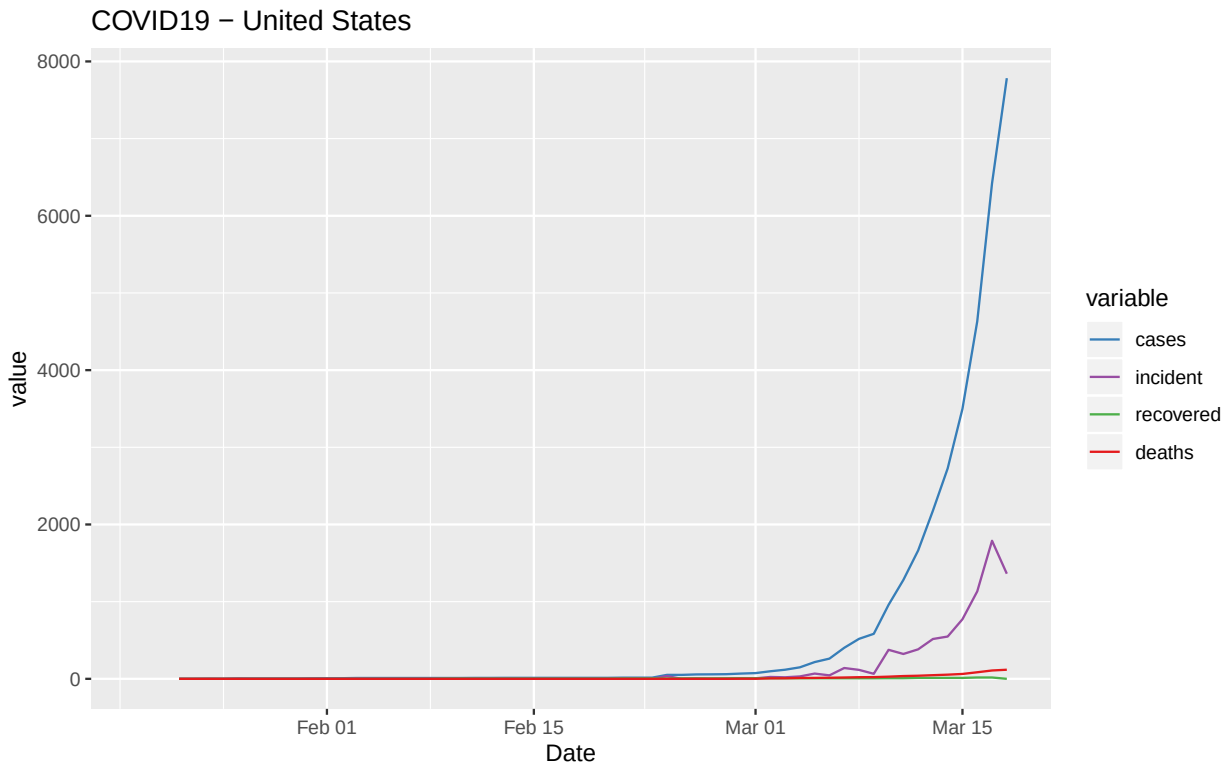

COVID19 − France
value
7500
5000
2500
0
Feb 01
Feb 15
Mar 01
Mar 15
Date
variable
cases
incident
recovered
deaths

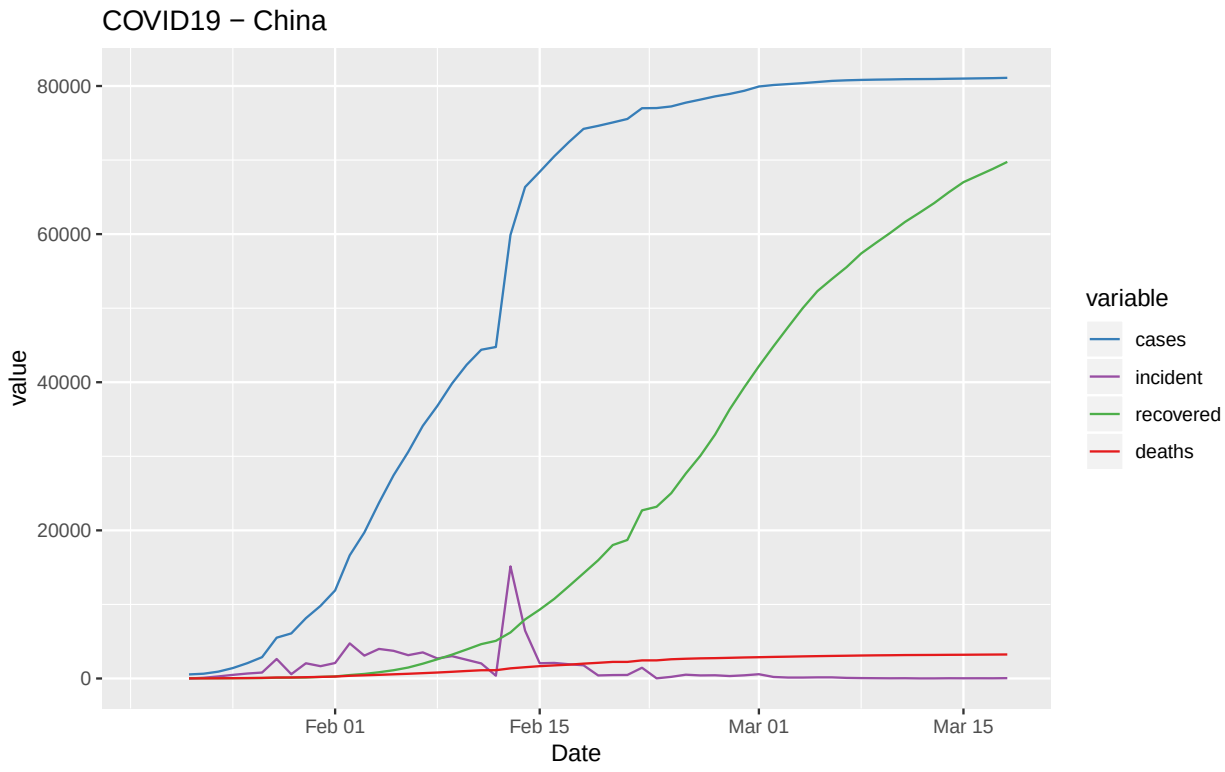

COVID19 − Austria
1500
1000
500
0
value
Feb 24
Mar 02
Mar 09
Mar 16
Date
variable
cases
incident
recovered
deaths


COVID19 − Slovakia
100
75
50
25
0
value
Mar 09
Mar 16
Date
variable
cases
incident
recovered
deaths


COVID19 − Poland
value
250
200
150
100
50
0
Mar 02
Mar 09
Mar 16
Date
variable
cases
incident
recovered
deaths


COVID19 − South Korea
value
8000
6000
4000
2000
0
Feb 01
Feb 15
Mar 01
Mar 15
Date
variable
cases
incident
recovered
deaths





Latest number of cases across countries

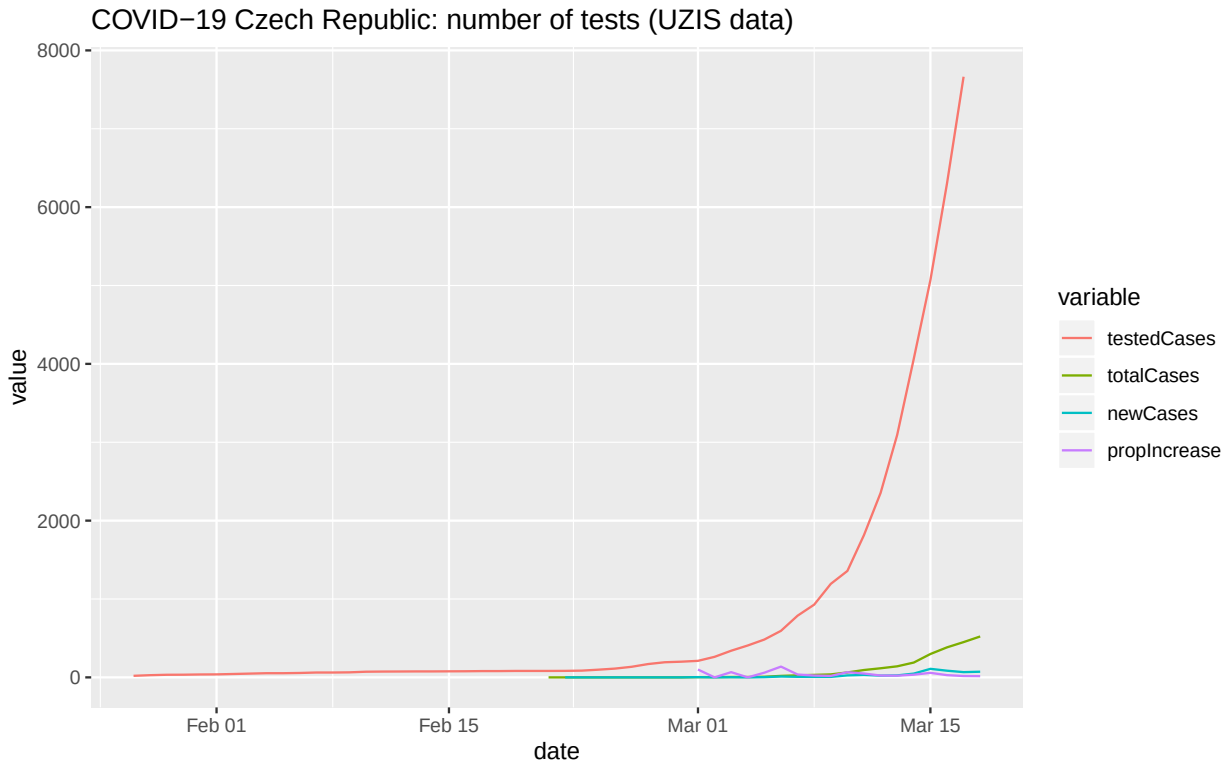
According to the Johns-Hopkins data

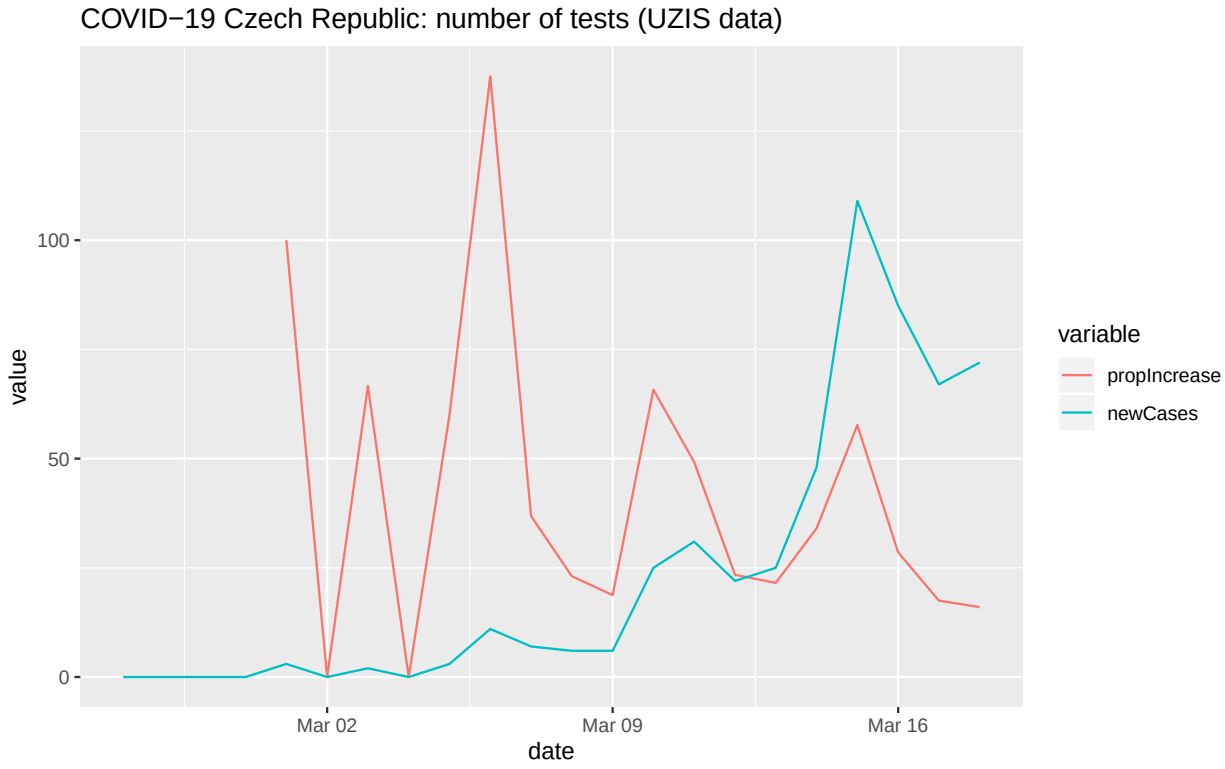
country	dateMax	count	deaths	recovered	deathRate
Austria	2020-03-18	1646	4	9	0.0024301
Czech Republic	2020-03-18	464	0	3	0.0000000
France	2020-03-18	9105	148	12	0.0162548
Germany	2020-03-18	12327	28	105	0.0022714
China	2020-03-18	81102	3241	69755	0.0399620
Iran	2020-03-18	17361	1135	5389	0.0653764
Italy	2020-03-18	35713	2978	4025	0.0833870
Poland	2020-03-18	251	5	13	0.0199203
Slovakia	2020-03-18	105	1	0	0.0095238
South Korea	2020-03-18	8413	84	1540	0.0099845
Spain	2020-03-18	13910	623	1081	0.0447879
United States	2020-03-18	7783	118	0	0.0151612

Plot and analyze Czech Rep. - CZ data

obtain latest czech data from MZ / apify https://api.apify.com/v2/key-value-stores/K373S4uCFR9W1K8ei/records/LATEST?disableRedirect=true (parsed from https://onemocneni-aktualne.mzcr.cz/covid-19)

exploratory plots: UZIS data





first case in CZ: 2020-03-01

Mean daily increase: 39.8%

The diseased count doubles on average every 2.07 days (49.6 hours).

Mean daily increase (last 5 days): 30.7%

The diseased count doubles on average every 2.59 days (62.1 hours).

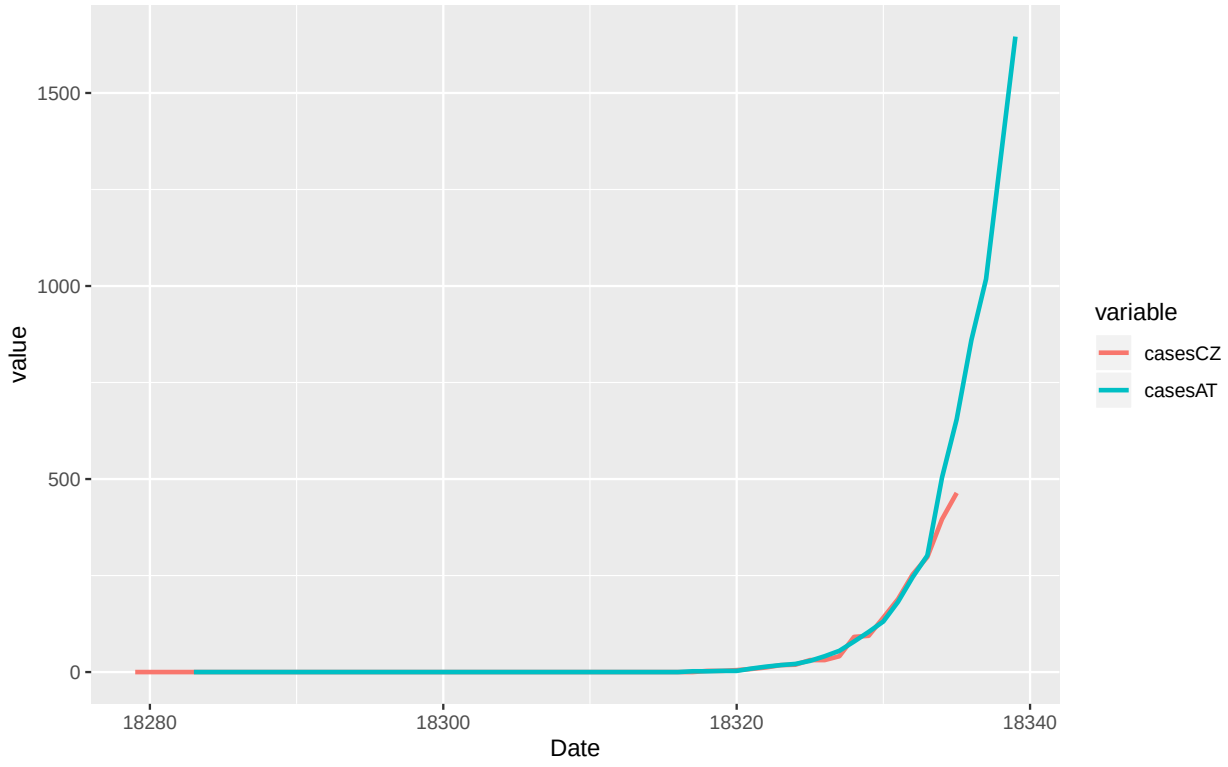
CZ data table

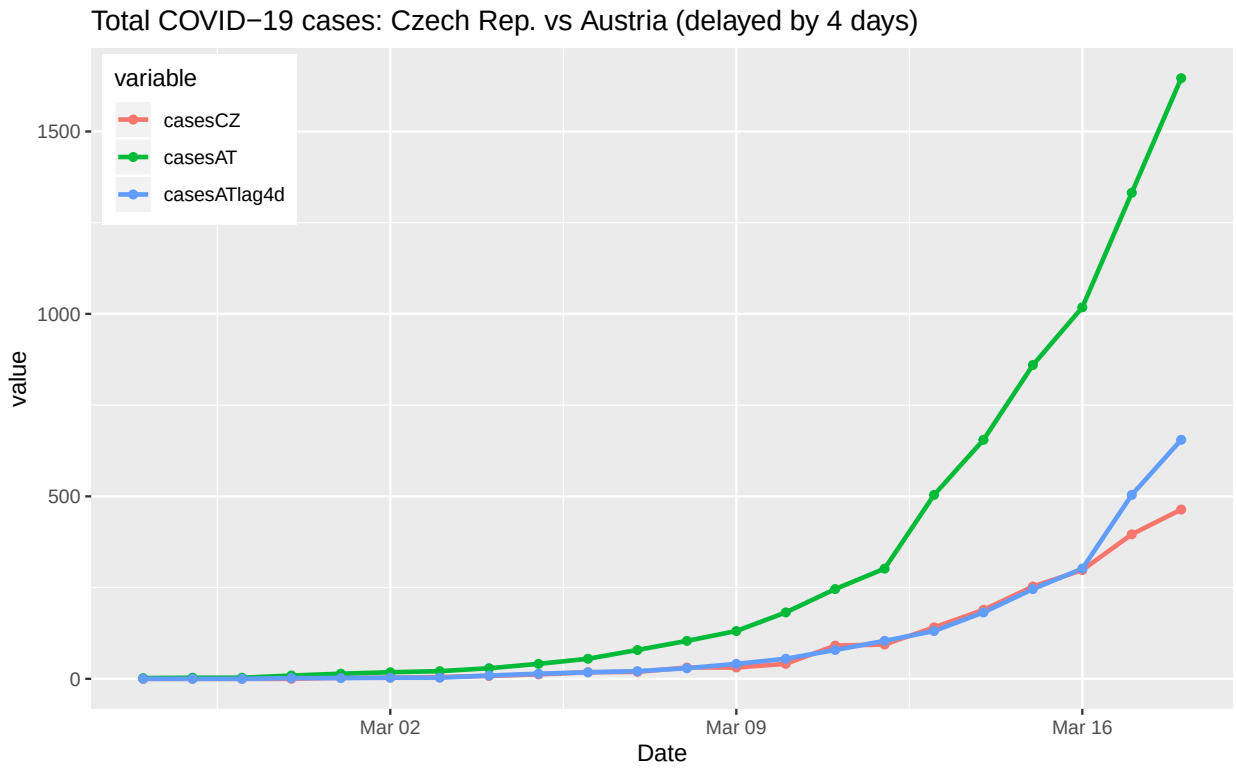
date	testedCases	totalCases	newCases	propIncrease
2020-01-27	20	NA	0	0.0000000
2020-01-28	28	NA	NA	NA
2020-01-29	33	NA	NA	NA
2020-01-30	34	NA	NA	NA
2020-01-31	37	NA	NA	NA
2020-02-01	38	NA	NA	NA
2020-02-02	43	NA	NA	NA
2020-02-03	48	NA	NA	NA
2020-02-04	53	NA	NA	NA
2020-02-05	53	NA	NA	NA
2020-02-06	56	NA	NA	NA
2020-02-07	62	NA	NA	NA
2020-02-08	62	NA	NA	NA
2020-02-09	64	NA	NA	NA
2020-02-10	72	NA	NA	NA
2020-02-11	74	NA	NA	NA
2020-02-12	75	NA	NA	NA
2020-02-13	76	NA	NA	NA
2020-02-14	76	NA	NA	NA
2020-02-15	77	NA	NA	NA
2020-02-16	78	NA	NA	NA
2020-02-17	80	NA	NA	NA
2020-02-18	80	NA	NA	NA
2020-02-19	82	NA	NA	NA
2020-02-20	82	NA	NA	NA
2020-02-21	82	0	NA	NA
2020-02-22	83	0	0	NaN
2020-02-23	86	0	0	NaN
2020-02-24	98	0	0	NaN
2020-02-25	112	0	0	NaN
2020-02-26	135	0	0	NaN
2020-02-27	170	0	0	NaN
2020-02-28	193	0	0	NaN
2020-02-29	200	0	0	NaN
2020-03-01	211	3	3	1.0000000
2020-03-02	262	3	0	0.0000000
2020-03-03	340	5	2	0.6666667
2020-03-04	407	5	0	0.0000000
2020-03-05	483	8	3	0.6000000
2020-03-06	594	19	11	1.3750000
2020-03-07	787	26	7	0.3684211
2020-03-08	928	32	6	0.2307692
2020-03-09	1193	38	6	0.1875000
2020-03-10	1358	63	25	0.6578947
2020-03-11	1816	94	31	0.4920635
2020-03-12	2353	116	22	0.2340426

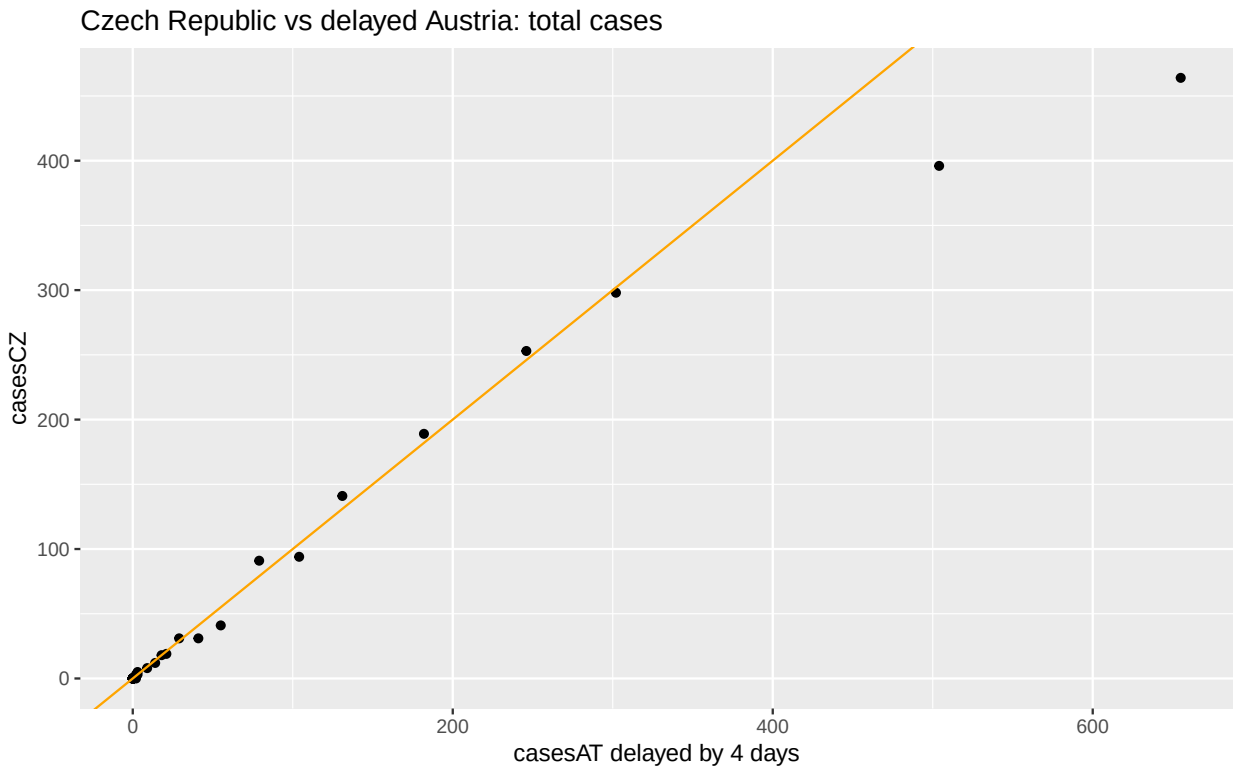
date	testedCases	totalCases	newCases	propIncrease
2020-03-13	3094	141	25	0.2155172
2020-03-14	4065	189	48	0.3404255
2020-03-15	5068	298	109	0.5767196
2020-03-16	6302	383	85	0.2852349
2020-03-17	7664	450	67	0.1749347
2020-03-18	NA	522	72	0.1600000

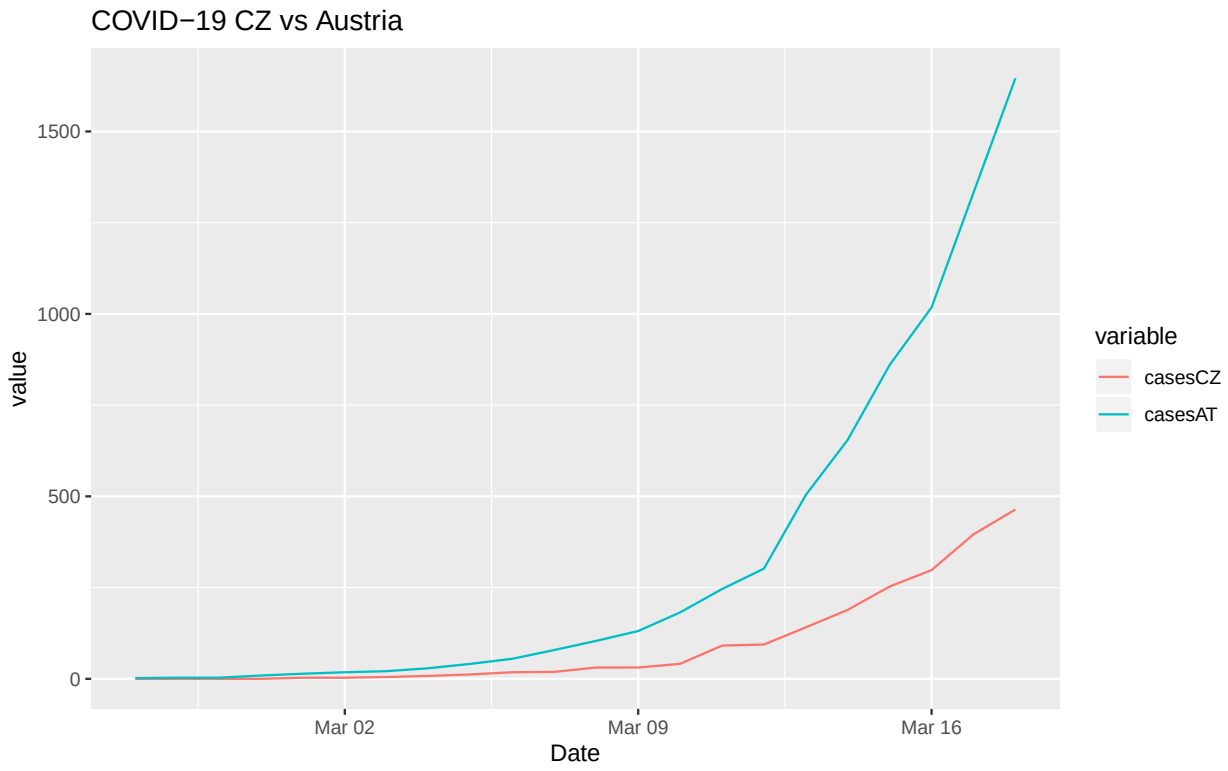
Austria vs CZ

AIM: estimate, by how many days the Czech Rep. is delayed behind other countries. First comparison: Austria




Total COVID−19 cases: Czech Rep. vs Austria (delayed by 4 days)
variable
casesCZ
casesAT
casesATlag4d
1500
1000
500
0
value
Mar 02
Mar 09
Mar 16
Date


Czech Republic vs delayed Austria: total cases
400
300
200
100
0
casesCZ
0
200
400
600
casesAT delayed by 4 days

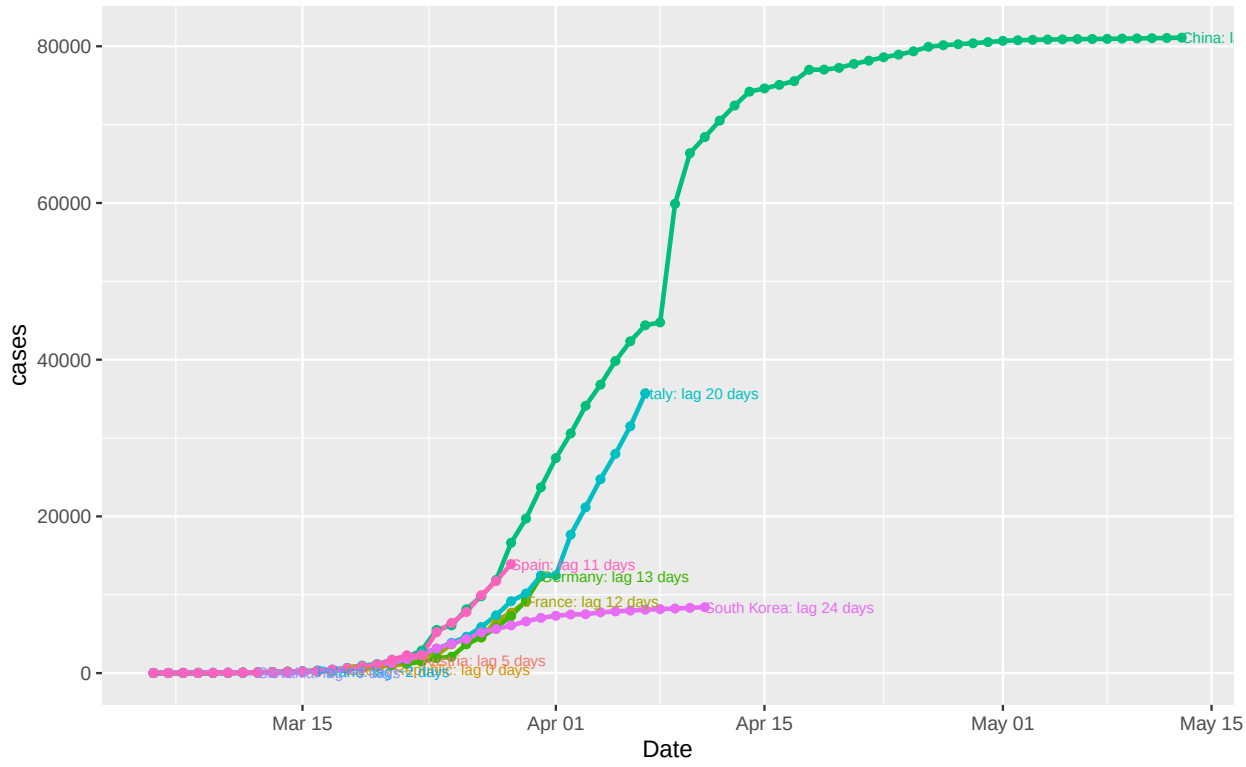


CZ vs others

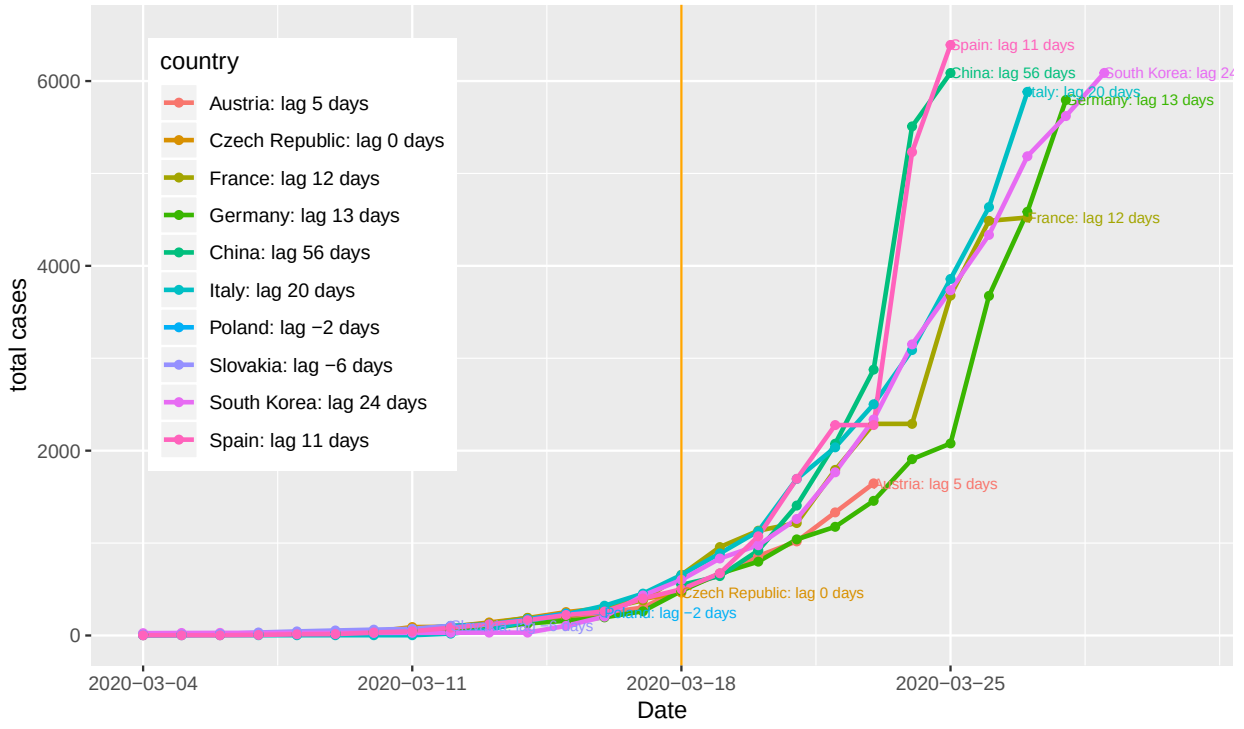
The lag selected by minimizing the RMS between total cases in CZ and total cases in the comparison countries. This compares the initial segment only and we can thus see, which scenarios may follow...

Country	CZ delayed by (days)
Austria	5
Italy	20
Germany	13
Spain	11
France	12
Slovakia	-6
Poland	-2
China	56
South Korea	24

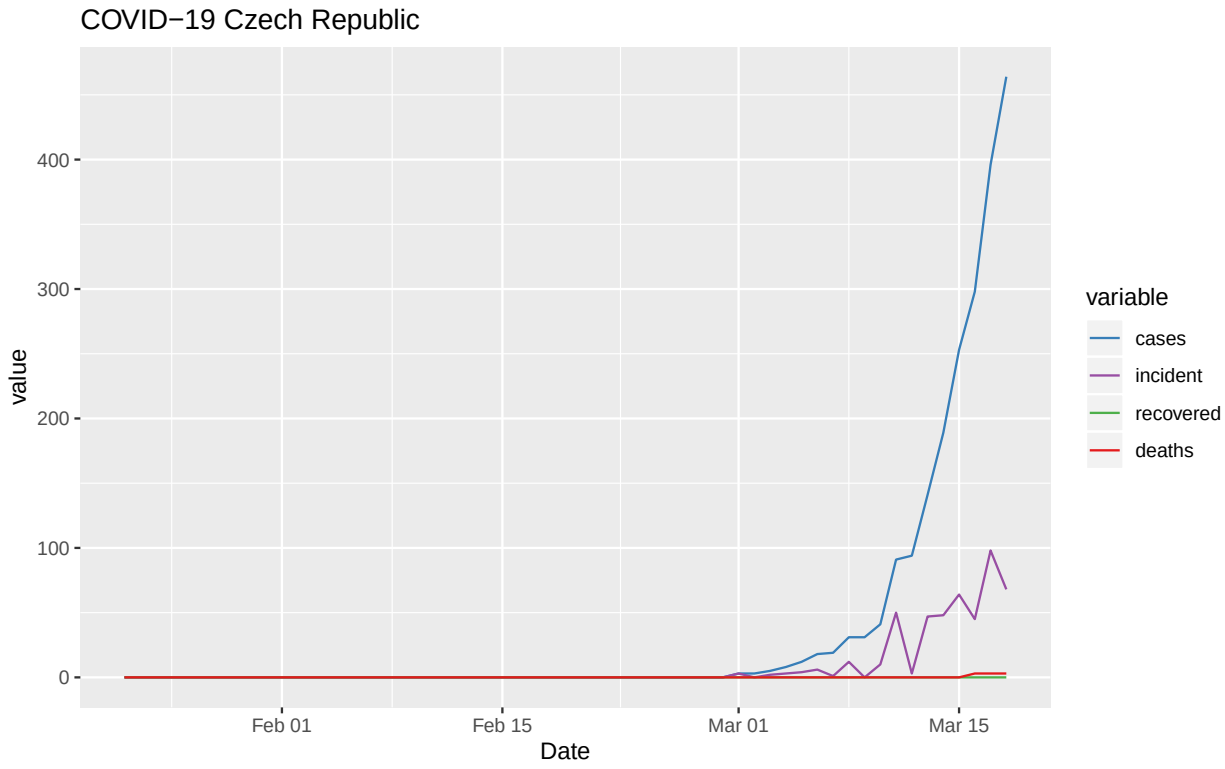
Note: tested integer delays in the range -14, 56 days



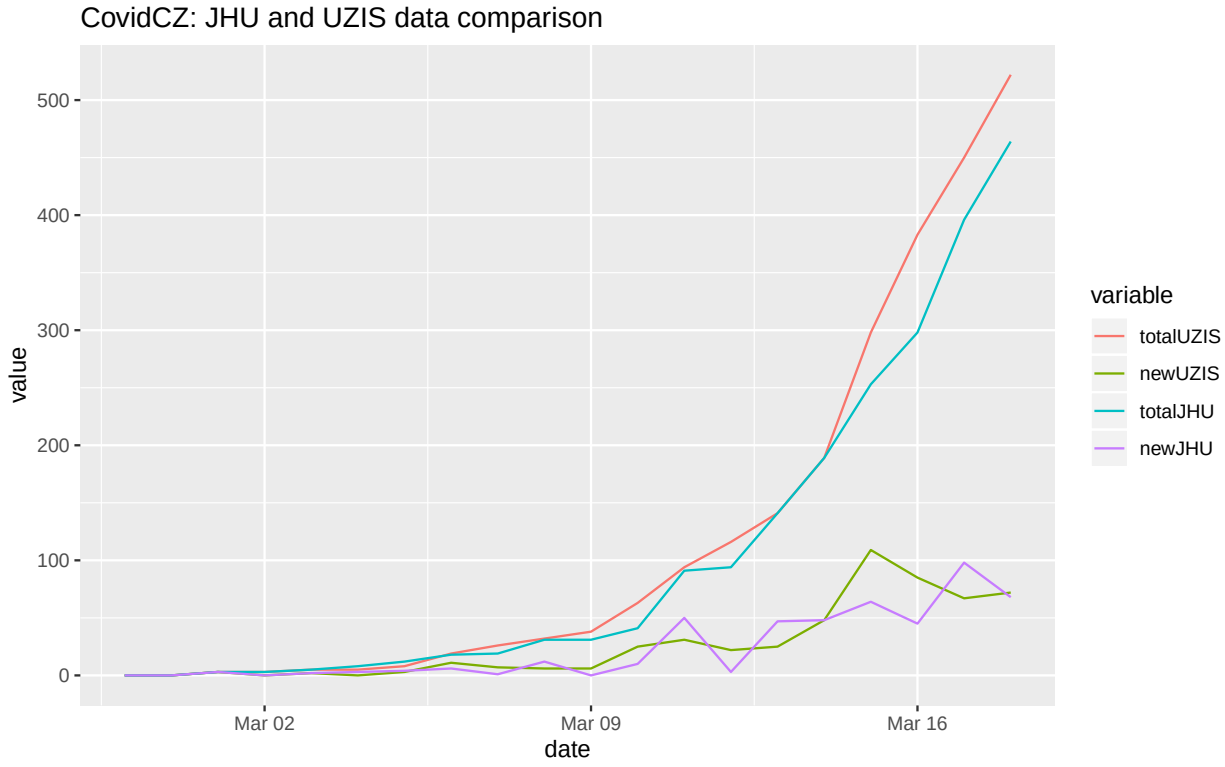
COVID19 trends: Czech Rep. vs lagged comparison countries



exploratory plots: JHU data



compare JHU and ÚZIS data



Epidemiological modelling - TBD

TBD according to: https://rviews.rstudio.com/2020/03/05/covid-19-epidemiology-with-r/ And https://timchurches.github.io/blog/posts/2020-03-10-modelling-the-effects-of-public-health-interventions-on-covid-19-transmission-part-1/

Store data locally for further reff