Analyse der COVID-19 Fallzahlen

Praxisprojekt

Regina Galambos, Lorenz Mihatsch

Projektpartner: André Klima

08. Mai 2020

Inhaltsangabe

COVID-19 Pandemie

1. COVID-19 ist ein Erkrankung, die durch das SARS-CoV-2 Virus ausgelöst wird.
2. Die Erkrankung ist erstmalig im Dezember 2019 in Wuhan (China) aufgetreten, der genaue Ursprung des Virus ist jedoch noch immer unbekannt.
3. Erster Erkrankungsfall in Deutschland am 28. Januar in Stockdorf.
4. Am 11. März wurde die ursprüngliche Epidemie (*Def.* örtlich begrenztes Auftreten einer ansteckenden Krankheit) von der WHO als Pandemie (*Def.* nicht mehr örtlich begrenztes Auftreten) eingestuft.
5. Am 22. März einigten sich Bund und Länder auf eine umfassende Beschränkung sozialer Kontakte.

COVID-19 Pandemie

1. COVID-19 ist ein Erkrankung, die durch das SARS-CoV-2 Virus ausgelöst wird.
2. Die Erkrankung ist erstmalig im Dezember 2019 in Wuhan (China) aufgetreten, der genaue Ursprung des Virus ist jedoch noch immer unbekannt.
3. Erster Erkrankungsfall in Deutschland am 28. Januar in Stockdorf.
4. Am 11. März wurde die ursprüngliche Epidemie (*Def.* örtlich begrenztes Auftreten einer ansteckenden Krankheit) von der WHO als Pandemie (*Def.* nicht mehr örtlich begrenztes Auftreten) eingestuft.
5. Am 22. März einigten sich Bund und Länder auf eine umfassende Beschränkung sozialer Kontakte.
6. **Wie haben sich die Zahl der Fälle und Todesfälle entwickelt?**

Datengrundlage

1. Anzahl an Fällen, Todesfällen und Genesenen
   - Datensatz des *Centers of Systems Science and Engineering* der Johns Hopkins University
   - Zusammentragung von Daten aus verschiedenen Quellen zu 213 Ländern. Bspw. *Italy Ministry of Health*.
   - Täglich von *RamiKrispin* auf GitHub aktualisiert und zu Verfügung gestellt. `https://github.com/RamiKrispin/coronavirus`

Datengrundlage

1. Anzahl an Fällen, Todesfällen und Genesenen
   - Datensatz des *Centers of Systems Science and Engineering* der Johns Hopkins University
   - Zusammentragung von Daten aus verschiedenen Quellen zu 213 Ländern. Bspw. *Italy Ministry of Health*.
   - Täglich von *RamiKrispin* auf GitHub aktualisiert und zu Verfügung gestellt. `https://github.com/RamiKrispin/coronavirus`
2. Demographie: Kontinentzugehörigkeit und Populationsdaten von 2018
   - Datenbank der Weltbank und der UN. Zugriff über das R-package *wbstat* und *JohnSnowLabs* über `https://datahub.io/JohnSnowLabs/country-and-continent-codes-list`.

Datengrundlage

1. Anzahl an Fällen, Todesfällen und Genesenen
   - Datensatz des *Centers of Systems Science and Engineering* der Johns Hopkins University
   - Zusammentragung von Daten aus verschiedenen Quellen zu 213 Ländern. Bspw. *Italy Ministry of Health*.
   - Täglich von *RamiKrispin* auf GitHub aktualisiert und zu Verfügung gestellt. `https://github.com/RamiKrispin/coronavirus`
2. Demographie: Kontinentzugehörigkeit und Populationsdaten von 2018
   - Datenbank der Weltbank und der UN. Zugriff über das R-package *wbstat* und *JohnSnowLabs* über `https://datahub.io/JohnSnowLabs/country-and-continent-codes-list`.
3. Politische Maßnahmen zur Eindämmung der Infektionsübertragung
   - *Government Response Tracker* Datesatz der *University of Oxford*.
   - Zusammentragung von Daten zu politische Maßnahmen der einzelnen Länder aus verschiedenen Quellen.
   - `https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker`

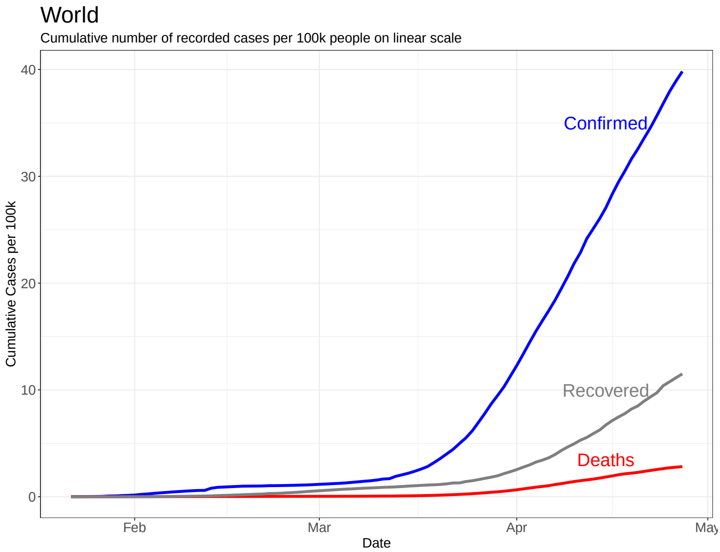
Anmerkungen zu den Daten

1. Es handelt sich “nur” um die aufgezeichneten Fälle und Todefälle. (max. untere Schranke)
2. Starke Unterschiede in der Aufzeichnungs- und Testpolitik einzelner Länder, was einen direkten Ländervergleich schwer möglich macht.
3. Die Kreuzfahrtschiffe Diamond Princess, Grand Princess und MS Zaandam wurden ausgeschlossen, da keinem Land eindeutig zuortbar.
4. Die Zahl der Fälle und Todesfälle beziehen sich meist auf 100.000 Personen.
5. Analyse endet am 27. April 2020.

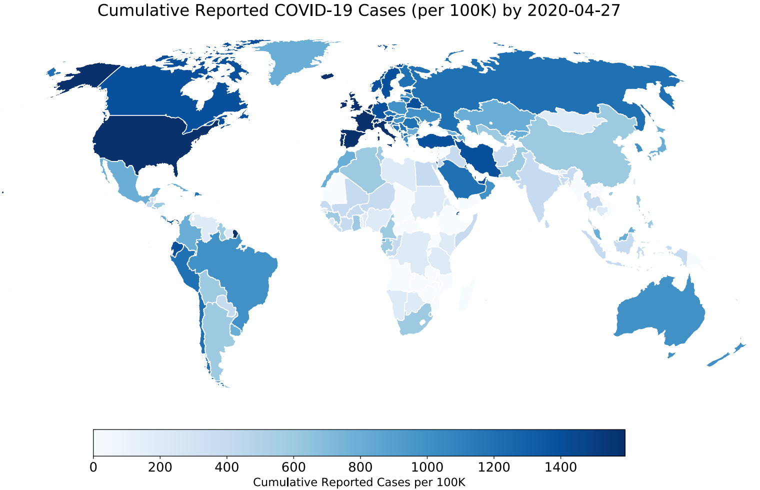
Anmerkungen zu den Daten

1. Es handelt sich "nur" um die aufgezeichneten Fälle und Todefälle. (max. untere Schranke)
2. Starke Unterschiede in der Aufzeichnungs- und Testpolitik einzelner Länder, was einen direkten Ländervergleich schwer möglich macht.
3. Die Kreuzfahrtschiffe Diamond Princess, Grand Princess und MS Zaandam wurden ausgeschlossen, da keinem Land eindeutig zuortbar.
4. Die Zahl der Fälle und Todesfälle beziehen sich meist auf 100.000 Personen.
5. Analyse endet am 27. April 2020.
6. Programmierung einer Web-Application: `https://github.com/mediasittich/covid19-dash-app`
7. Abbildungen sind auf Englisch da sie sich an ein ggf. internationales Publikum richten.

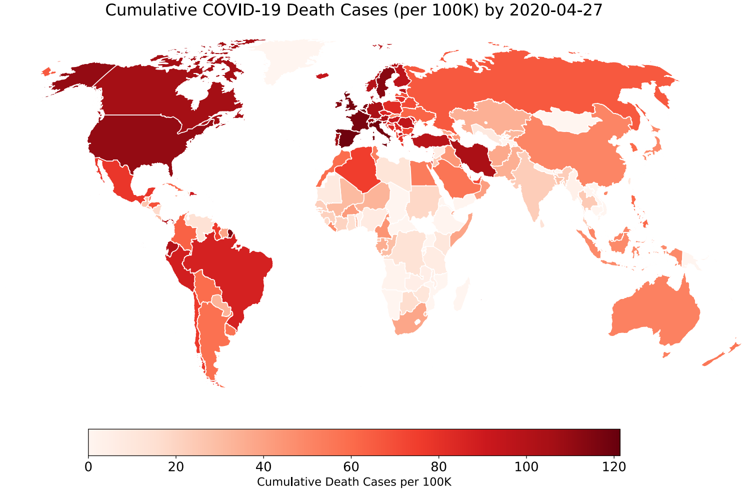
COVID-19 weltweit



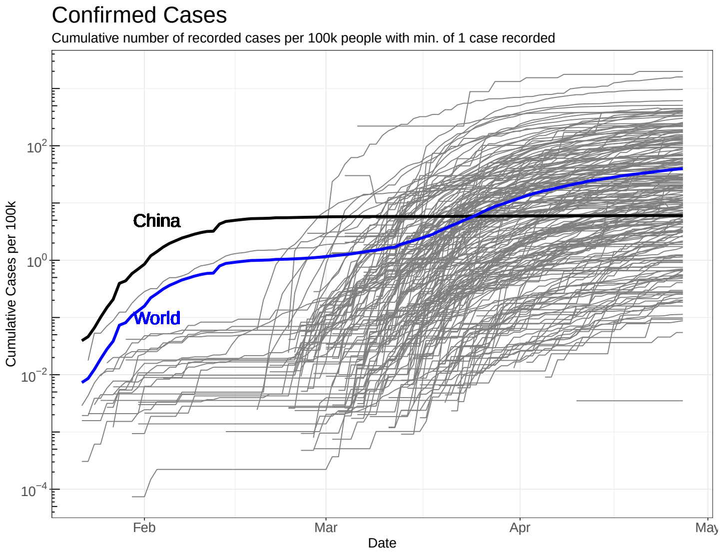
COVID-19 weltweit



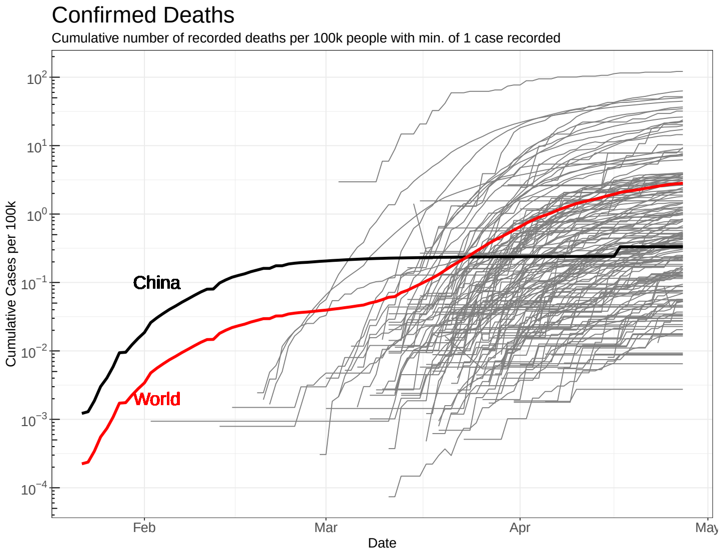
COVID-19 weltweit



COVID-19 weltweit bestätigte Fälle



COVID-19 weltweit bestätigte Todesfälle



Zwischenergebnis

- Aktuell (27. April 2020) sind weltweit etwa 40 Fälle und 3 Todesfälle pro 100k Einwohner gemeldet.
- Durch logarithmische Darstellung sind zwei Infektionswellen erkennbar. Die erste Infektionswelle ist durch China dominiert.
- Geografisch besonders Europa und die USA betroffen.
- Für die meisten afrikanischen Staaten sind kaum Fälle gemeldet.

Wiederholung: Wachstumsfaktor und geometrisches Mittel

Definition 1 *Wachstumsfaktor*
Sei $C_0, C_1, C_2, ...$ eine Zeitreihe von kumulativen Fallzahlen zu den Zeitpunkten $0, 1, ..., n$. Dann ist für $i = 1, ..., n$ der i-te Wachstumsfaktor x_1 gegeben durch

$$x_i = \frac{C_i}{C_{i-1}}.$$

Dadruch ergeben sich die kum. Fallzahlen zum Zeitpunkt n durch $C_n = C_0 \cdot x_1 \cdot x_2 \cdot ... \cdot x_n$.

Beispiel: Deutschland $C_{01.03.} = 130; C_{02.03.} = 159 \Rightarrow x_{02.03.} = \frac{159}{130} = 1.22$

Wiederholung: Wachstumsfaktor und geometrisches Mittel

Definition 2 *Geometrisches Mittel*
Das geometrische Mittel zu den Wachstumsfaktoren $x_1, x_2, ..., x_n$ ist gegeben durch

$$\bar{x}_{geom} = (x_1 \cdot x_2 \cdot ... \cdot x_n)^{1/n}.$$

Daraus ergibt sich $C_n = C_0 \cdot x_1 \cdot x_2 \cdot ... \cdot x_n = C_0 \cdot (\bar{x}_{geom})^n$.

Wiederholung: Wachstumsfaktor und geometrisches Mittel

Definition 2 *Geometrisches Mittel*
Das geometrische Mittel zu den Wachstumsfaktoren $x_1, x_2, ..., x_n$ ist gegeben durch

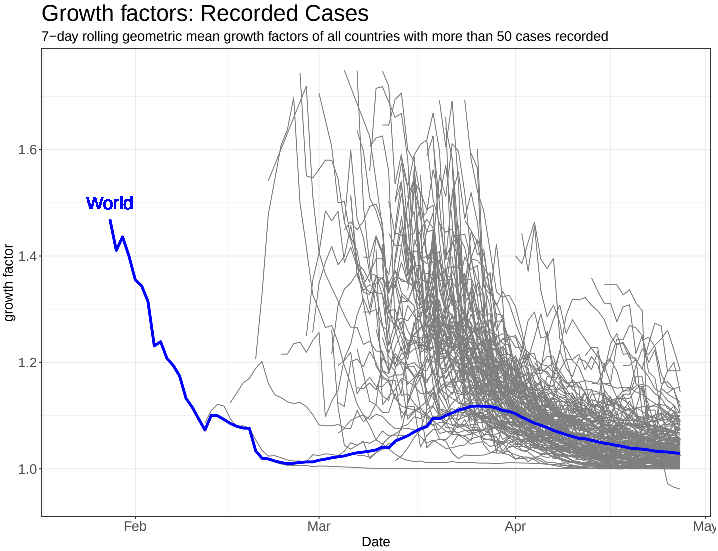
$$\bar{x}_{geom} = (x_1 \cdot x_2 \cdot ... \cdot x_n)^{1/n}.$$

Daraus ergibt sich $C_n = C_0 \cdot x_1 \cdot x_2 \cdot ... \cdot x_n = C_0 \cdot (\bar{x}_{geom})^n$.

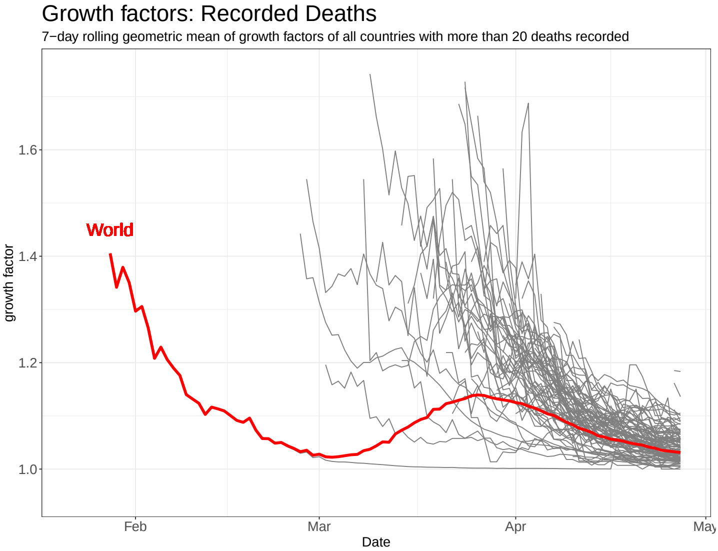
Wir betrachten im Folgenden den *rolling geometric mean* der vergangenen 7 Tage. Dazu berechnen wir für jeden Zeitpunkt i

$$\bar{x}_{i,geom} = (x_i \cdot x_{i-1} \cdot x_{i-2} \cdot ... \cdot x_{i-6})^{1/7}.$$

Wachstumsfaktoren: Bestätigte Fälle



Wachstumsfaktoren: Todesfälle



Verdopplungszeit

Ausgehend von einem exponentiellen Wachstum der Form $C_n = C_0 \cdot (\bar{x}_{n,geom})^n$ ergibt sie die "momentane" Verdopplunszeit dt_i der Fallzahlen durch

$$dt_i = \frac{ln(2)}{ln(\bar{x}_{i,geom})}.$$

Verdopplungszeit

Ausgehend von einem exponentiellen Wachstum der Form $C_n = C_0 \cdot (\bar{x}_{n,geom})^n$ ergibt sie die "momentane" Verdopplunszeit dt_i der Fallzahlen durch

$$dt_i = \frac{ln(2)}{ln(\bar{x}_{i,geom})}.$$

Beispiel: Erinnerung: Wachstumsfaktor von Deutschland am 02. März 2020

$$x_{02.03.} = 1.22 \Rightarrow dt_{02.03.} = \frac{ln(2)}{ln(1.22)} = 3.49$$

Verdopplungszeit

Ausgehend von einem exponentiellen Wachstum der Form $C_n = C_0 \cdot (\bar{x}_{n,geom})^n$ ergibt sie die "momentane" Verdopplunszeit dt_i der Fallzahlen durch

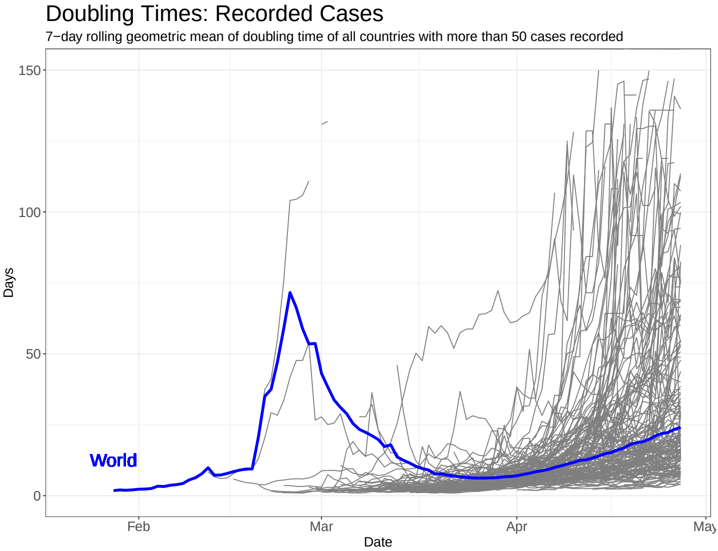
$$dt_i = \frac{ln(2)}{ln(\bar{x}_{i,geom})}.$$

Beispiel: Erinnerung: Wachstumsfaktor von Deutschland am 02. März 2020

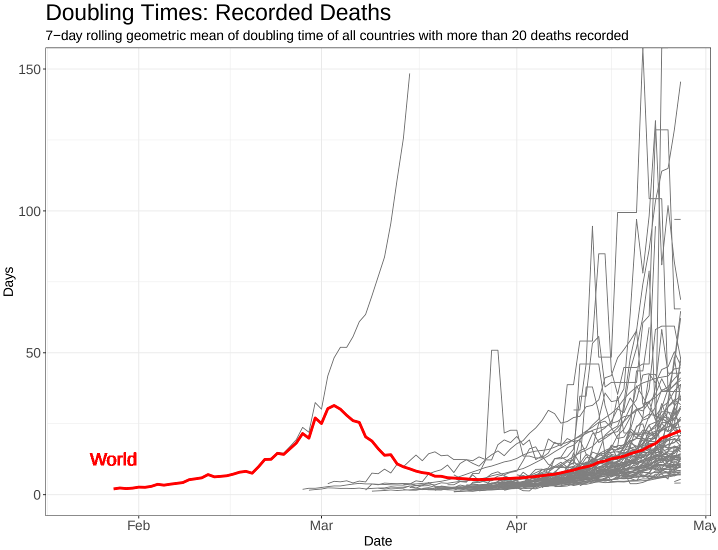
$$x_{02.03.} = 1.22 \Rightarrow dt_{02.03.} = \frac{ln(2)}{ln(1.22)} = 3.49$$

Bei einem Wachstum mit Faktor 1.22 würde sich die Anzahl an Fällen also alle 3.49 Tage verdoppeln.

Verdopplungszeit: Bestätigte Fälle



Verdopplungszeit: Todesfälle



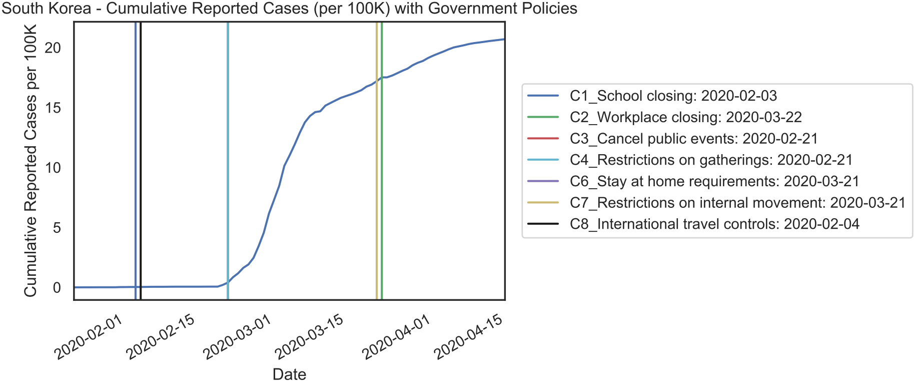
Zwischenergebnis

- Unter der Annahme eines exponentiellen Wachstums lassen sich kumulativen Fällen und Todesfällen über die Berechnung von Wachstumsfaktoren und Verdopplungszeiten direkt miteinander vergleichen.
- Aktuell (27. April 2020) besteht ein Wachstumsfaktor der kumulativen Fälle von etwa 1.03, was einer Verdopplungszeit von etwa 24 Tagen entspricht.
- Das lokale Minimum der Wachstumsfaktoren (bzw. lokales Maximum der Verdopplungszeit) der weltweiten Fälle liegt etwa eine Woche vor dem lokalen Minimun der Wachstumsfaktoren der weltweiten Todesfälle.
- Mitte-Ende März tritt sowohl bei den Wachstumsfaktoren von Fällen und Todesfällen ein lokales Maximum der Wachstumsfaktoren (bzw. lokales Minimum der Verdopplungszeit) auf.

Diskussion

- Daten geben maximal eine untere Schranke der Zahl der Fälle und Todesfälle an.
- Weltweit zwei Infektionswellen, wobei die Erste von China dominiert ist.
- Unter Annahme eines exponentiellen Wachstums zeigt sich das Maximum der zweiten Infektionswelle Mitt-Ende März.
- Weltweit sind die Wachstumsfaktoren in den letzen Tagen rückläufig.

Ausblick: Infektionsmaßnahmen



Ende

Vielen herzlichen Dank für Ihre Aufmerksamkeit!

Herleitung der Verdopplungszeit

Ausgehend von einem exponentiellen Wachstum der Form $C_n = C_0 \cdot (\bar{x}_{n,geom})^n$ ergibt sie die "momentane" Verdopplunszeit dt_i der Fallzahlen durch

$$dt_i = \frac{ln(2)}{ln(\bar{x}_{i,geom})}.$$

Herleitung:

$$\begin{aligned} C_i \cdot (\bar{x}_{i,geom})^{dt_i} = 2 \cdot C_i &\iff (\bar{x}_{i,geom})^{dt_i} = 2 \\ &\iff dt_i = \frac{ln(2)}{ln(\bar{x}_{i,geom})}. \end{aligned}$$