

Analyse der COVID-19 Fallzahlen

Praxisprojekt

Regina Galambos, Lorenz Mihatsch



Projektpartner: André Klima
08. Mai 2020

Inhaltsangabe

- 1 Einführung und Fragestellung
- 2 Daten
- 3 Kumulativen Anzahl an Fällen und Todesfällen weltweit
- 4 Wachstumsfaktoren der kumulativen Fälle und Todesfälle
- 5 Einfluss von Social Distancing auf Infektionsübertragung
- 6 Diskussion

COVID-19 Pandemie

- ① COVID-19 ist eine 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 eine 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 **Welchen Einfluss haben Einschränkungen sozialer Kontakte auf die Verbreitung der COVID-19 Erkrankung?**

Datengrundlage

① Anzahl an Fällen, Todesfällen und Genesenen

- Datensatz des *Centers of Systems Science and Engineering* der John Hopkins University
- Zusammentragung von Daten aus verschiedenen Quellen zu — 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 John Hopkins University
 - Zusammentragung von Daten aus verschiedenen Quellen zu — 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 John Hopkins University
 - Zusammentragung von Daten aus verschiedenen Quellen zu — 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 pol. Maßnahmen der einzelnen Länder aus verschiedenen Quellen.
 - <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>

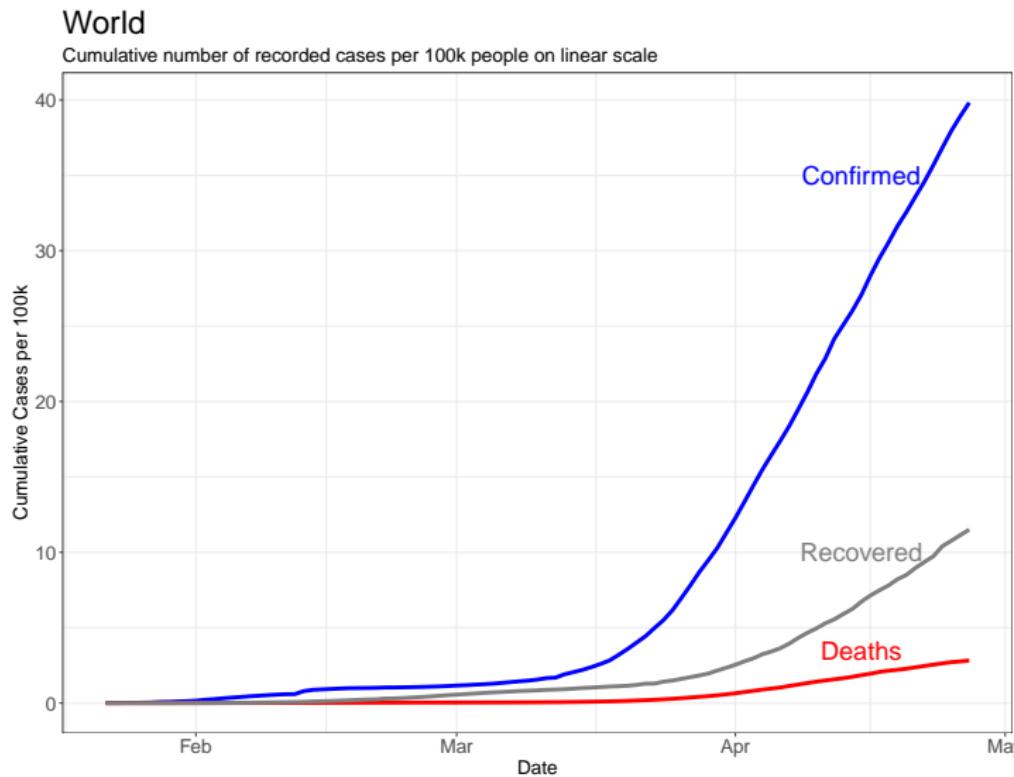
Anmerkungen zu den Daten

- ① Es handelt sich “nur” um die aufgezeichneten Fälle und Todesfälle.
(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, keine Land eindeutig zuortenbar.
- ④ Die Zahlen der Fälle und Todesfälle beziehen sich meist auf 100.000 Personen.
- ⑤ Analyse endet am 27. April 2020.

Anmerkungen zu den Daten

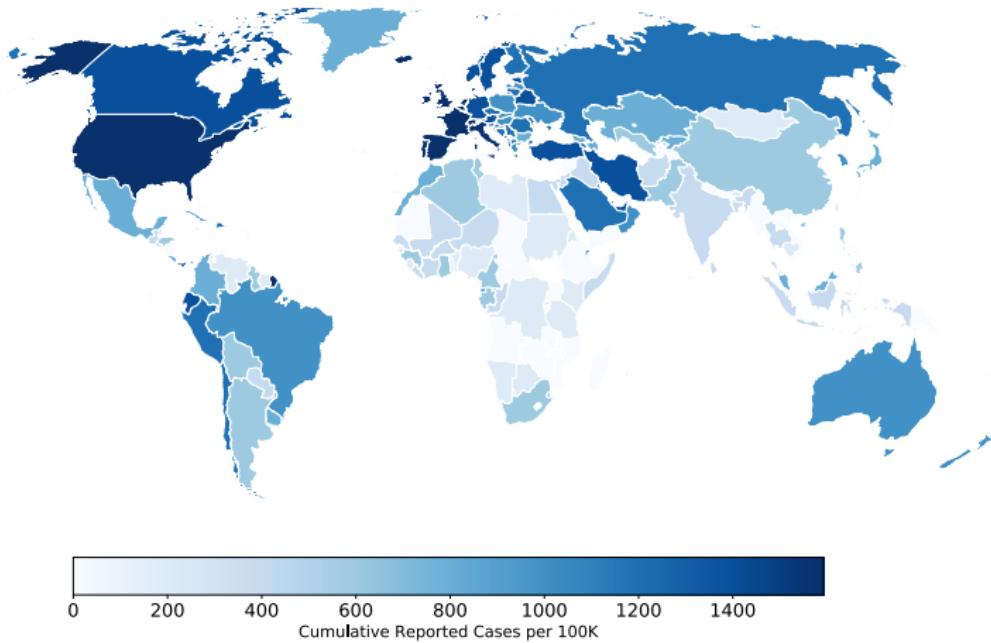
- ① Es handelt sich "nur" um die aufgezeichneten Fälle und Todesfälle.
(max. untere Schranke)
- ② Starke Unterschiede in der Aufzeichnungs- und Testpolitik einzelner Länder, was einen direkten Ländervergleich schwer möglich macht.
- ③ Die Kreuzfahrtschiffe Diamond Princess, Grand Princess und MS Zaandam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

COVID-19 weltweit



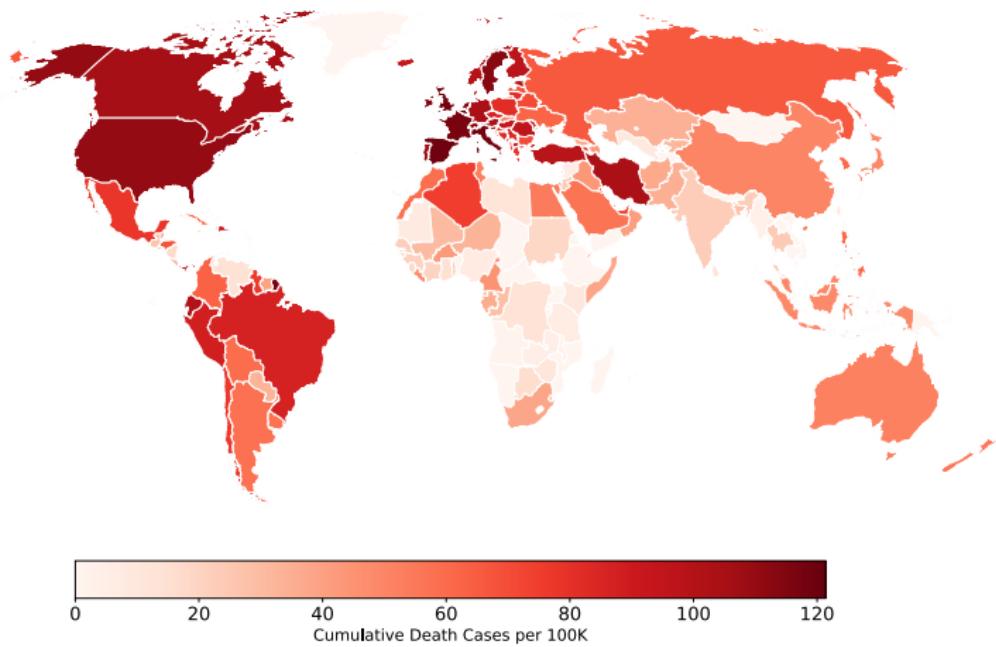
COVID-19 weltweit

Cumulative Reported COVID-19 Cases (per 100K) by 2020-04-27

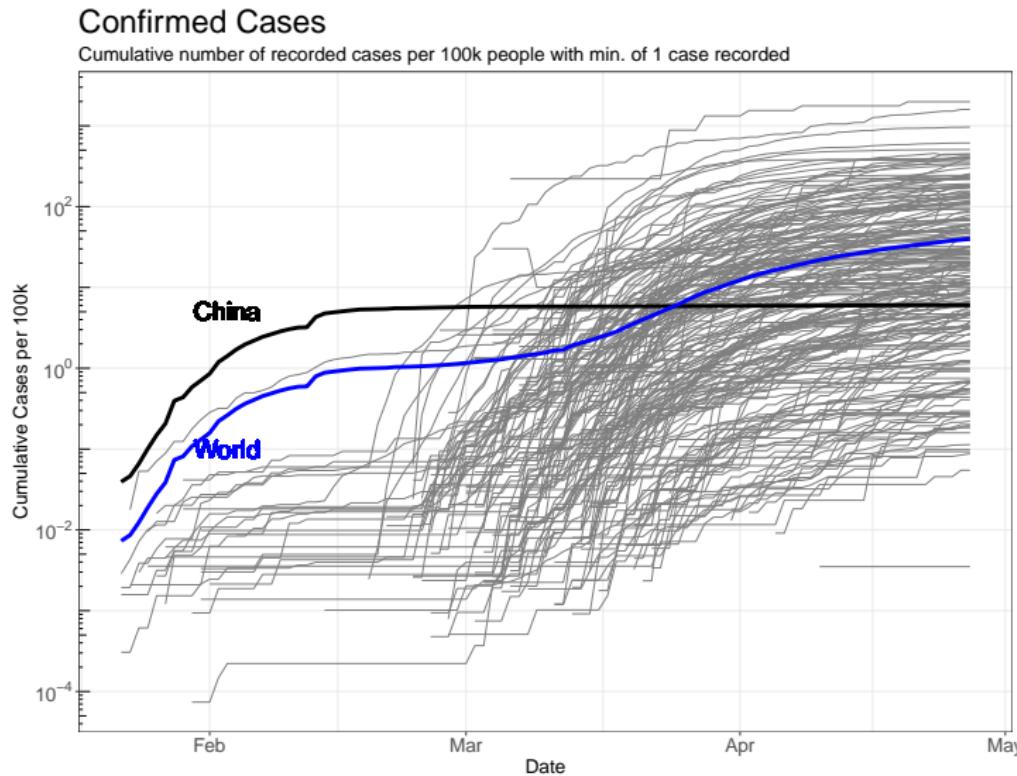


COVID-19 weltweit

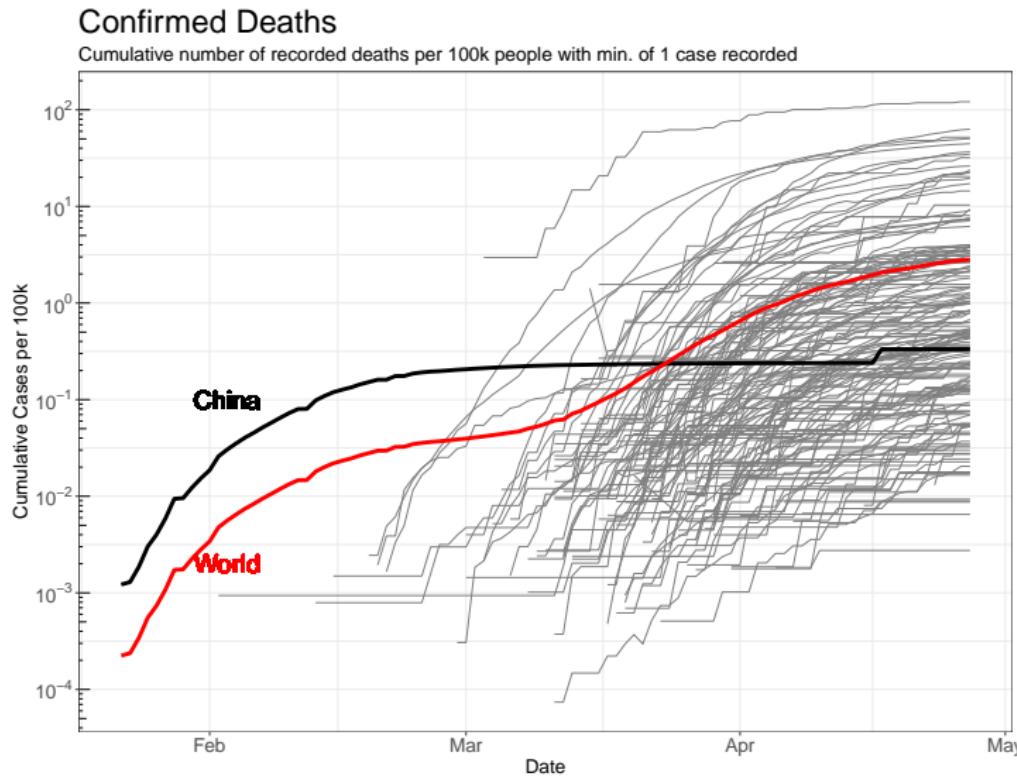
Cumulative COVID-19 Death Cases (per 100K) by 2020-04-27



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.
- Geographisch besonderd Europa und die USA bestroffen.
- Für die meisten afrikanischen Staaten sind kaum Fälle gemeldet.

Wiederholung: Wachstumsfaktor und geometrisches Mittel

Definition 1 Wachstumsfaktor

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

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

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

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, \dots, x_n ist gegeben durch

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

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

Wiederholung: Wachstumsfaktor und geometrisches Mittel

Definition 2 Geometrisches Mittel

Das geometrische Mittel zu den Wachstumsfaktoren x_1, x_2, \dots, x_n ist gegeben durch

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

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

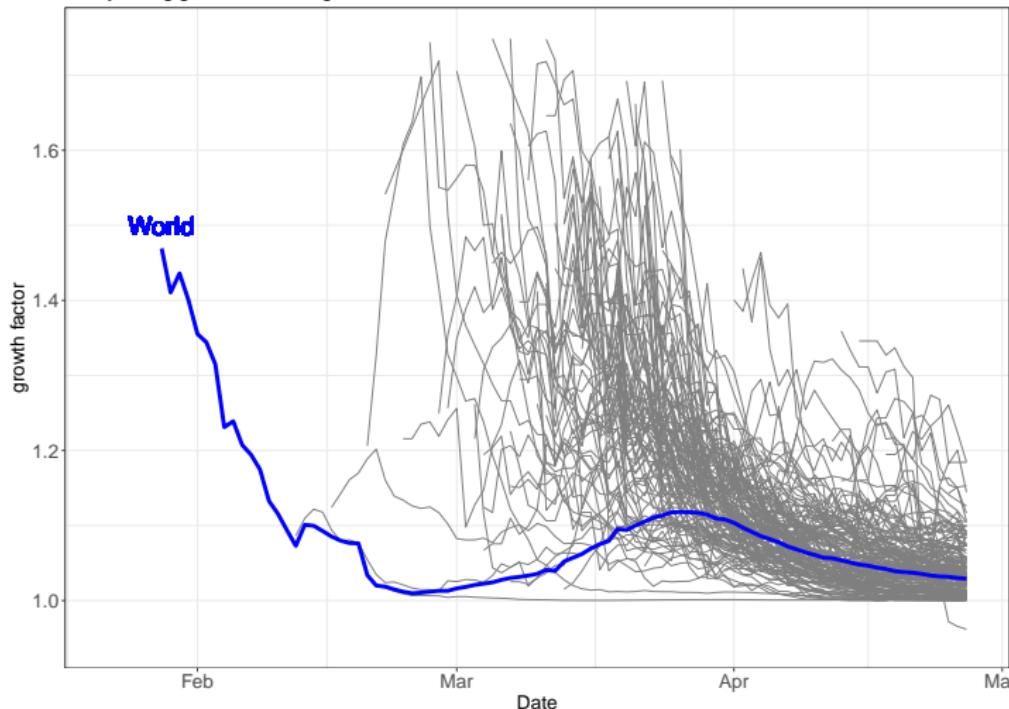
Wir betrachten im Folgenden den *rolling geometric mean* der vergangenen 7 Tage. Dazu berechnen wir für jeden Zeitpunkt i

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

Wachstumsfaktoren: Bestätigte Fälle

Growth factors: Recorded Cases

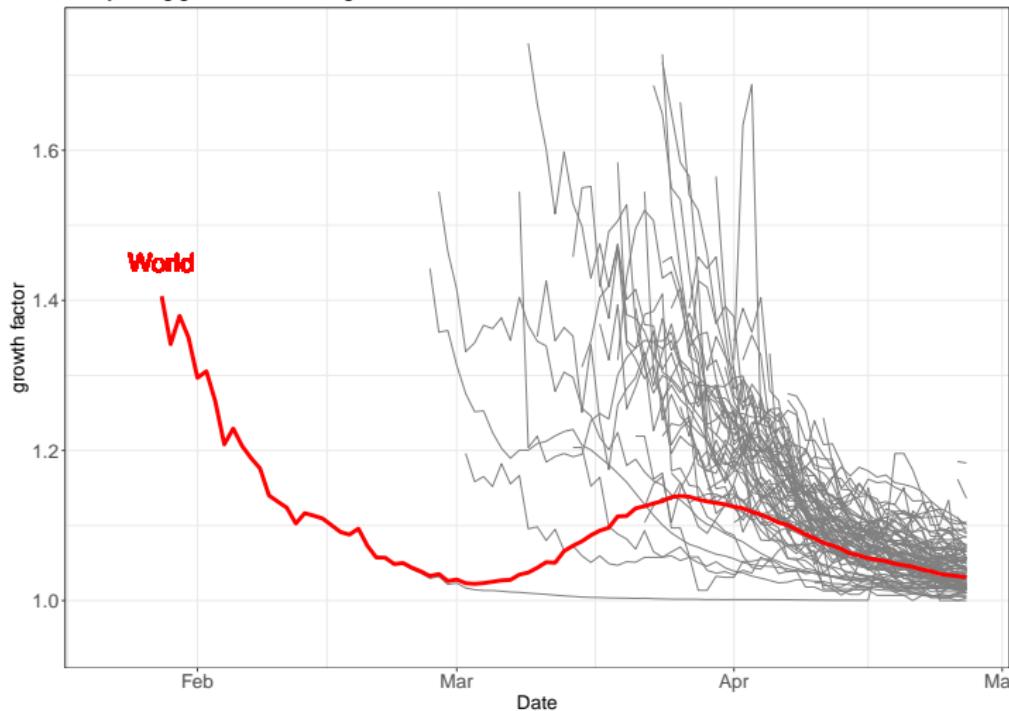
7-day rolling geometric mean growth factors of all countries with more than 50 cases recorded



Wachstumsfaktoren: Todesfälle

Growth factors: Recorded Deaths

7-day rolling geometric mean of growth factors of all countries with more than 20 deaths recorded



Verdopplungszeit

Ausgehend von einem exponentiellen Wachstum der Form

$C_n = C_0 \cdot (\bar{x}_{n,geom})^n$ ergibt sie die "momentane" Verdopplungszeit 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" Verdopplungszeit 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" Verdopplungszeit dt_i der Fallzahlen durch

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

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

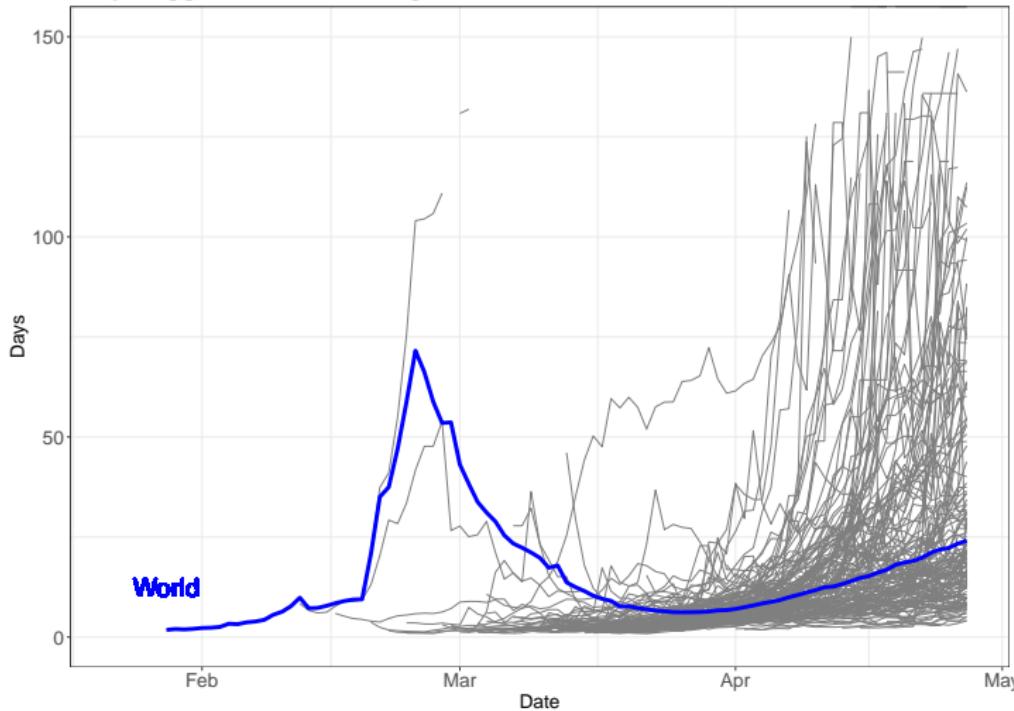
$$x_{02.03.} = 1.22 \Rightarrow dt_{02.03.} = \frac{\ln(2)}{\ln(1.22)} = 3.49$$

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

Verdopplungszeit: Bestätigte Fälle

Doubling Times: Recorded Cases

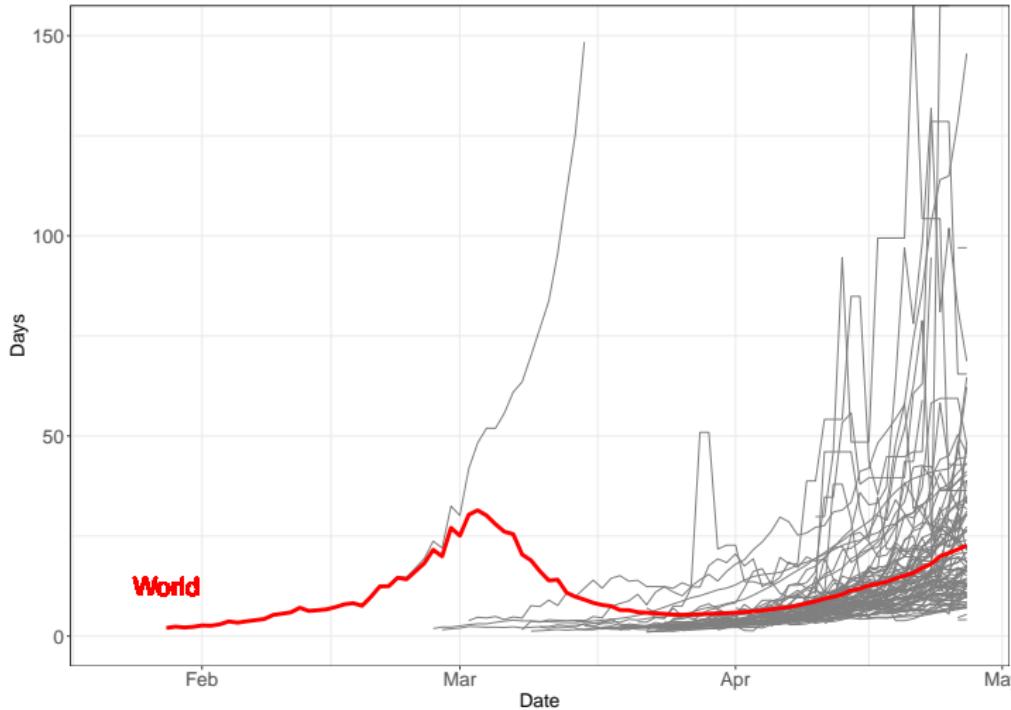
7-day rolling geometric mean of doubling time of all countries with more than 50 cases recorded



Verdopplungszeit: Todesfälle

Doubling Times: Recorded Deaths

7-day rolling geometric mean of doubling time of all countries with more than 20 deaths recorded



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.
- Der lokale Minimum der Wachstumsfaktoren (bzw. lokales Maximum der Verdopplungszeit) der weltweiten Fälle liegt etwa eine Woche vor dem lokalen Minimum 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.

Infektionsmaßnahmen

S6_Restrictions on internal movement in selected countries



Diskussion

- Die Berechnung des *geometrischen Mittels der Wachstumsfaktoren* und der *Verdopplungszeiten* beruhen auf der Annahme eines exponentielle Wachstums. Zulässigkeit?
- Starke Unterschiede in der Strenge der Ausgangsbeschränkungen einzelner Länder.
- Verschiedene Maßnahmen machen Gruppierung nur schwer möglich.
- Erlaubt keine Schlüsse zur Ursache und Wirkung der abfallenden Wachstumsfaktoren.

Ende

Vielen Herzlichen Dank für Ihre Aufmerksamkeit!

Referenzen

- R-Packages
- RamiKrispin
- Oxford
- JHU
- UN Weltbank

Herleitung der Verdopplungszeit

Ausgehend von einem exponentiellen Wachstum der Form

$C_n = C_0 \cdot (\bar{x}_{n,geom})^n$ ergibt sie die "momentane" Verdopplungszeit dt_i der Fallzahlen durch

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

Herleitung:

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