**Raw open source data**

Alle data komt van <https://github.com/mzelst/covid-19>, behalve rioolwaterdata van <https://data.rivm.nl/covid-19/>

Lcps.csv

* Data IC inflow vanaf 17-10-2020 beschikbaar
* Raw variables
  + IC Bedden COVID: Het aantal IC-bedden bezet door patiënten met COVID-19;
  + IC Bedden Non-COVID: Het aantal IC-bedden bezet door patiënten zonder COVID-19;
  + IC Nieuwe Opnames COVID: Het aantal patiënten met COVID-19 dat in 24 uur nieuw is opgenomen op de IC;
  + Kliniek bedden: Het aantal kliniekbedden bezet door patiënten met COVID-19.
  + Kliniek Nieuwe Opnames COVID: Het aantal patiënten met COVID-19 dat in 24 uur nieuw is opgenomen in de kliniek;
* Artificial variables
  + Totaal bezetting: Aantal kliniek en IC bedden bezet door patienten met COVID-19.
  + IC Opnames 7d: 7 daagse gemiddelde van IC Nieuwe Opnames COVID
  + Kliniek Opnames 7d: 7 daagse gemiddelde van Kliniek Nieuwe Opnames COVID
  + Totaal Opnames: Kliniek Nieuwe Opnames COVID + IC Nieuwe Opnames COVID
  + Totaal Opnames 7d: 7 daagse gemiddelde van Totaal Opnames
  + Totaal IC: Totaal aantal patienten op de IC
  + IC Opnames 14d: 14 daagse gemiddelde van IC Nieuwe Opnames COVID
  + Kliniek Opnames 14d: 14 daagse gemiddelde van Kliniek Nieuwe Opnames COVID
  + OMT Check IC: ???
  + OMT Check Kliniek: ???

Rivm\_tests.csv

* RIVM dashboard: Aantal (positieve) testen per dag voor heel Nederland
* Variables
  + Tested total
  + Infected
  + Infected percentage
  + Tested total 7d average
  + Positive rate 3d average
  + Positive rate 7d average

Ggd\_tests\_per\_agegroup.csv

* GGD: positieve testen per dag per GGD per leeftijdsgroep
  + Welke open source deze data van mzelst github precies vandaan komt is mij onbekend
  + Data moet geaggregeerd worden per leeftijdsgroep

Rivm\_reproduction\_number.csv

* RIVM R-getal: <https://coronadashboard.rijksoverheid.nl/verantwoording>
  + Voor de schatting van dit reproductiegetal gebruikt het RIVM het aantal gemelde positieve COVID-19 testuitslagen per dag.
  + Tot 12 juni werd het reproductiegetal berekend op basis van COVID-19 ziekenhuisopnames omdat er tot dan toe minder getest werd.
* Variables
  + R lower bound
  + R average
  + R upper bound

Rivm\_infectious\_people.csv

* RIVM Besmettelijke mensen (data tot 2 mei):
  + Zie definitie op <https://coronadashboard.rijksoverheid.nl/verantwoording>
  + Vanaf 13 oktober wordt dit cijfer berekend op basis van de data uit de tweede ronde van het Pienter Corona onderzoek en het aantal ziekenhuisopnames.
* Variables
  + Prevalence lower bound
  + Prevalence average
  + Prevalence upper bound
  + Groei besmettelijken
  + Groei besmettelijken 7d average

Rivm\_gedrag.csv

* RIVM Gedrag (data tot 20 april beschikbaar):
  + <https://coronadashboard.rijksoverheid.nl/verantwoording#gedrag>
  + <https://data.rivm.nl/geonetwork/srv/dut/catalog.search#/metadata/8a72d78a-fcf8-4882-b0ab-cd594961a267>
* Data transformaties
  + Data moet van long naar wide format
  + Data per veiligheidsregio kan weg
  + Data elke 3 weken geactualiseerd

Vaccines.csv

* Vaccinatiedata
  + Welke open source deze data van mzelst github precies vandaan komt is mij onbekend
  + In het originele databestand op mzelst github is elke publicatie van vaccinatiedata onder elkaar geplakt in excel. Ik heb alle niet relevante rijen verwijderd en alleen de meest recente info behouden
* Variables
  + Estimated vaccines carehomes
  + Administered vaccines GGD
  + Administered vaccines hospital
  + Estimated administered vaccines total (vanaf 24 jan wijkt dit af van administered)
  + Administered vaccines total
  + Administered vaccines doctors

Rivm\_rioolwaterdata.csv

* Van 30-03-2020 tot 06-09-2020 werden RNA deeltjes per ml bijgehouden. Vanaf 07-09-2020 tot nu werd RNA deeltjes per 100.000 inwoners bijgehouden.
* Er zijn 319 meetlocaties in Nederland in onze data. Het aantal metingen loopt van 1tot 91. Het gemiddelde is 51.3887.
* Data is hierdoor niet erg betrouwbaar. Er zit geen patroon in. Met een 3- en zeker7-daags gemiddelde is dit wel zo
* Mogelijk is misschien om de data per ROAZ regio in te delen. Dan heb je meermetingen per gebied. Ook een zevendaags gemiddelde per ROAZ-regio is een optie
* Het gesommeerde RNA per 100000 per dag kunnen we delen door het aantal metingen om voor het aantal metingen te controleren
* In deze link zijn het aantal inwoners per zuiveringsinstallatie te zien:https://www.cbs.nl/nl-nl/maatwerk/2021/06/inwoners-per-rioolwaterzuiveringsinstallatie-1-1-2021

Onbruikbare data:

* Data-dashboairds/positieve-testen-per-week-per-leeftijdsgroep
  + RIVM dashboard: Aantal positieve testen per leeftijdsgroep per week
  + Data is niet geupdate na week 11 2021 in mzelst github, dus ONBRUIKBAAR
* NICE data over ziekenhuis en IC inflow: verschilt significant van LCPS data. Verschil zit in data op patientniveau (NICE) vs data op geaggregeerd bezettingsniveau (LCPS). Zie <https://lcps.nu/veelgestelde-vragen/>