

Week4.Datasets

Daniel Vogel

2/14/2020

R Markdown

Source: <https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>

Content

2019 Novel Coronavirus (2019-nCoV) is a virus (more specifically, a coronavirus) identified as the cause of an outbreak of respiratory illness first detected in Wuhan, China. Early on, many of the patients in the outbreak in Wuhan, China reportedly had some link to a large seafood and animal market, suggesting animal-to-person spread. However, a growing number of patients reportedly have not had exposure to animal markets, indicating person-to-person spread is occurring. At this time, it's unclear how easily or sustainably this virus is spreading between people - CDC

This dataset has daily level information on the number of affected cases, deaths and recovery from 2019 novel coronavirus. Please note that this is a time series data and so the number of cases on any given day is the cumulative number.

The data is available from 22 Jan, 2020. Column Description

2019_ncov_data.csv

Sno - Serial number Date - Date and time of the observation in MM/DD/YYYY HH:MM:SS Province / State - Province or state of the observation (Could be empty when missing) Country - Country of observation Last Update - Time in UTC at which the row is updated for the given province or country. (Not standardised currently. So please clean them before using it) Confirmed - Number of confirmed cases Deaths - Number of deaths Recovered - Number of recovered cases

Acknowledgements

Johns Hopkins university has made the data available in google sheets format here. Sincere thanks to them.

Thanks to WHO, CDC, NHC and DXY for making the data available in first place.

Some insights could be

Changes in number of affected cases over time

Change in cases over time at country level

Latest number of affected cases

Summary of datasets

```
summary(ncov_df)
```

##	Sno	Date	Province/State	Country
##	Min. : 1.0	Length:1419	Length:1419	Length:1419
##	1st Qu.: 355.5	Class :character	Class :character	Class :character
##	Median : 710.0	Mode :character	Mode :character	Mode :character

```
## Mean      : 710.0
## 3rd Qu.:1064.5
## Max.      :1419.0
## Last Update      Confirmed      Deaths      Recovered
## Length:1419      Min.       :    0.0   Min.       :   0.000   Min.       :    0.00
## Class :character  1st Qu.:    2.0   1st Qu.:   0.000   1st Qu.:    0.00
## Mode  :character  Median :   11.0   Median :   0.000   Median :    0.00
##                      Mean      : 353.1   Mean      :   7.838   Mean      :   24.92
##                      3rd Qu.:   95.0   3rd Qu.:   0.000   3rd Qu.:    4.00
##                      Max.      :51986.0   Max.      :1426.000   Max.      :4131.00
```

```
summary(ncov_deaths_df)
```

```
## Province/State      Country/Region      Lat      Long
## Length:72           Length:72           Min.    : -37.81   Min.    : -123.12
## Class :character     Class :character     1st Qu.: 23.80     1st Qu.: 17.12
## Mode  :character     Mode  :character     Median : 34.05     Median : 106.00
##                      Mean      : 30.24     Mean      : 64.66
##                      3rd Qu.: 41.16     3rd Qu.: 117.25
##                      Max.      : 61.92     Max.      : 153.03
##
## 1/21/20 22:00 1/22/20 12:00 1/23/20 12:00 1/24/20 0:00 1/24/20 12:00
## Mode:logical Mode:logical Mode:logical Min.      : 1.000   Min.      : 1.000
## NA's:72      NA's:72      NA's:72      1st Qu.: 1.000   1st Qu.: 1.000
##                      Median : 1.000   Median : 1.000
##                      Mean      : 8.667   Mean      : 8.667
##                      3rd Qu.:12.500   3rd Qu.:12.500
##                      Max.      :24.000   Max.      :24.000
##                      NA's      :69      NA's      :69
## 1/25/20 0:00 1/25/20 12:00 1/25/20 22:00 1/26/20 11:00 1/26/20 23:00
## Min.      : 1.00   Min.      : 1.0   Min.      : 1.00   Min.      : 1.0   Min.      : 1
## 1st Qu.: 1.00   1st Qu.: 1.0   1st Qu.: 1.00   1st Qu.: 1.0   1st Qu.: 1
## Median : 1.00   Median : 1.0   Median : 1.00   Median : 1.0   Median : 1
## Mean      :11.33   Mean      :14.0   Mean      :13.75   Mean      :11.2   Mean      :16
## 3rd Qu.:16.50   3rd Qu.:20.5   3rd Qu.:13.75   3rd Qu.: 1.0   3rd Qu.: 1
## Max.      :32.00   Max.      :40.0   Max.      :52.00   Max.      :52.0   Max.      :76
## NA's      :69     NA's      :69     NA's      :68     NA's      :67     NA's      :67
## 1/27/20 9:00 1/27/20 19:00 1/27/20 20:30 1/28/20 13:00
## Min.      : 1.0   Min.      : 1.00   Min.      : 1.00   Min.      : 1.00
## 1st Qu.: 1.0   1st Qu.: 1.00   1st Qu.: 1.00   1st Qu.: 1.00
## Median : 1.0   Median : 1.00   Median : 1.00   Median : 1.00
## Mean      :13.5   Mean      :11.71   Mean      : 15.14   Mean      : 15.14
## 3rd Qu.: 1.0   3rd Qu.: 1.00   3rd Qu.: 1.00   3rd Qu.: 1.00
## Max.      :76.0   Max.      :76.00   Max.      :100.00   Max.      :100.00
## NA's      :66     NA's      :65     NA's      :65     NA's      :65
## 1/28/20 18:00 1/28/20 23:00 1/29/20 13:30 1/29/20 14:30
## Min.      : 1.00   Min.      : 1.00   Min.      : 1.00   Min.      : 1.00
## 1st Qu.: 1.00   1st Qu.: 1.00   1st Qu.: 1.00   1st Qu.: 1.00
## Median : 1.00   Median : 1.00   Median : 1.00   Median : 1.00
## Mean      : 18.71   Mean      : 18.86   Mean      : 18.86   Mean      : 16.62
## 3rd Qu.: 1.00   3rd Qu.: 1.50   3rd Qu.: 1.50   3rd Qu.: 1.25
## Max.      :125.00   Max.      :125.00   Max.      :125.00   Max.      :125.00
## NA's      :65     NA's      :65     NA's      :65     NA's      :64
## 1/29/20 21:00 1/30/20 11:00 1/31/20 14:00 02/01/20 10:00
## Min.      : 1.00   Min.      : 1.00   Min.      : 1.00   Min.      : 1.00
```

```
## 1st Qu.: 1.00 1st Qu.: 1.00 1st Qu.: 1.00 1st Qu.: 1.00
## Median : 1.00 Median : 1.00 Median : 1.00 Median : 1.00
## Mean : 21.25 Mean : 21.38 Mean : 26.62 Mean : 28.78
## 3rd Qu.: 1.25 3rd Qu.: 2.00 3rd Qu.: 2.00 3rd Qu.: 2.00
## Max. :162.00 Max. :162.00 Max. :204.00 Max. :249.00
## NA's :64 NA's :64 NA's :64 NA's :63
## 02/02/20 21:00 02/03/20 21:00 02/04/20 9:40 02/04/20 22:00 02/05/20 9:00
## Min. : 1.0 Min. : 1.0 Min. : 1.0 Min. : 1.00 Min. : 1
## 1st Qu.: 1.0 1st Qu.: 1.0 1st Qu.: 1.0 1st Qu.: 1.00 1st Qu.: 1
## Median : 1.0 Median : 1.0 Median : 1.0 Median : 1.00 Median : 1
## Mean : 36.2 Mean : 42.6 Mean : 42.6 Mean : 44.73 Mean : 38
## 3rd Qu.: 2.0 3rd Qu.: 2.0 3rd Qu.: 2.0 3rd Qu.: 2.00 3rd Qu.: 2
## Max. :350.0 Max. :414.0 Max. :414.0 Max. :479.00 Max. :479
## NA's :62 NA's :62 NA's :62 NA's :61 NA's :59
## 02/05/20 23:00 02/06/20 9:00 02/06/20 14:20 02/07/20 20:13
## Min. : 1.00 Min. : 1.00 Min. : 1.00 Min. : 1.00
## 1st Qu.: 1.00 1st Qu.: 1.00 1st Qu.: 1.00 1st Qu.: 1.00
## Median : 1.00 Median : 1.00 Median : 1.00 Median : 1.00
## Mean : 43.46 Mean : 43.46 Mean : 43.46 Mean : 42.53
## 3rd Qu.: 2.00 3rd Qu.: 2.00 3rd Qu.: 2.00 3rd Qu.: 2.00
## Max. :549.00 Max. :549.00 Max. :549.00 Max. :618.00
## NA's :59 NA's :59 NA's :59 NA's :57
## 02/07/20 22:50 02/08/20 10:24 02/08/20 23:04 02/09/20 10:30
## Min. : 1.00 Min. : 0.00 Min. : 0.00 Min. : 0.00
## 1st Qu.: 1.00 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00
## Median : 1.00 Median : 0.00 Median : 0.00 Median : 0.00
## Mean : 45.25 Mean : 10.07 Mean : 11.29 Mean : 11.31
## 3rd Qu.: 2.00 3rd Qu.: 0.00 3rd Qu.: 1.00 3rd Qu.: 1.00
## Max. :699.00 Max. :699.00 Max. :780.00 Max. :780.00
## NA's :56
## 02/09/20 23:20
## Min. : 0.00
## 1st Qu.: 0.00
## Median : 0.00
## Mean : 12.64
## 3rd Qu.: 1.00
## Max. :871.00
##
```

```
summary(ncov_confirmed_df)
```

```
## Province/State Country/Region Lat Long
## Length:72 Length:72 Min. : -37.81 Min. : -123.12
## Class :character Class :character 1st Qu.: 23.80 1st Qu.: 17.12
## Mode :character Mode :character Median : 34.05 Median : 106.00
## Mean : 30.24 Mean : 64.66
## 3rd Qu.: 41.16 3rd Qu.: 117.25
## Max. : 61.92 Max. : 153.03
##
## 1/21/20 22:00 1/22/20 12:00 1/23/20 12:00 1/24/20 0:00
## Min. : 1.00 Min. : 1.00 Min. : 1.00 Min. : 1.00
## 1st Qu.: 1.00 1st Qu.: 1.00 1st Qu.: 2.00 1st Qu.: 2.00
## Median : 2.00 Median : 2.00 Median : 3.00 Median : 4.00
## Mean : 20.62 Mean : 19.14 Mean : 17.68 Mean : 23.18
## 3rd Qu.: 6.00 3rd Qu.: 5.00 3rd Qu.: 7.00 3rd Qu.: 12.00
```

##	Max.	:270.00	Max.	:444.00	Max.	:444.00	Max.	:549.00
##	NA's	:56	NA's	:43	NA's	:35	NA's	:34
##	1/24/20	12:00	1/25/20	0:00	1/25/20	12:00	1/25/20	22:00
##	Min.	: 1.00	Min.	: 1.00	Min.	: 1.00	Min.	: 1.00
##	1st Qu.	: 2.00	1st Qu.	: 3.00	1st Qu.	: 3.00	1st Qu.	: 3.00
##	Median	: 4.50	Median	: 5.50	Median	: 8.00	Median	: 11.00
##	Mean	: 23.52	Mean	: 32.21	Mean	: 33.35	Mean	: 46.84
##	3rd Qu.	: 15.00	3rd Qu.	: 22.50	3rd Qu.	: 25.00	3rd Qu.	: 37.50
##	Max.	:549.00	Max.	:729.00	Max.	:761.00	Max.	:1052.00
##	NA's	:32	NA's	:30	NA's	:29	NA's	:29
##	1/26/20	11:00	1/26/20	23:00	1/27/20	9:00	1/27/20	19:00
##	Min.	: 1.00	Min.	: 1.00	Min.	: 1.00	Min.	: 1.00
##	1st Qu.	: 3.00	1st Qu.	: 4.00	1st Qu.	: 4.00	1st Qu.	: 4.00
##	Median	: 8.00	Median	: 11.00	Median	: 9.50	Median	: 8.00
##	Mean	: 43.22	Mean	: 57.02	Mean	: 57.72	Mean	: 57.37
##	3rd Qu.	: 36.00	3rd Qu.	: 47.00	3rd Qu.	: 47.75	3rd Qu.	: 50.00
##	Max.	:1058.00	Max.	:1423.00	Max.	:1423.00	Max.	:1423.00
##	NA's	:23	NA's	:23	NA's	:22	NA's	:21
##	1/27/20	20:30	1/28/20	13:00	1/28/20	18:00	1/28/20	23:00
##	Min.	: 1.00	Min.	: 1.00	Min.	: 1.0	Min.	: 1.0
##	1st Qu.	: 3.75	1st Qu.	: 3.00	1st Qu.	: 4.0	1st Qu.	: 4.0
##	Median	: 8.50	Median	: 11.00	Median	: 11.0	Median	: 13.0
##	Mean	: 86.02	Mean	: 88.49	Mean	: 105.2	Mean	: 114.3
##	3rd Qu.	: 60.75	3rd Qu.	: 66.00	3rd Qu.	: 66.0	3rd Qu.	: 80.0
##	Max.	:2714.00	Max.	:2714.00	Max.	:3554.0	Max.	:3554.0
##	NA's	:20	NA's	:19	NA's	:19	NA's	:19
##	1/29/20	13:30	1/29/20	14:30	1/29/20	21:00	1/30/20	11:00
##	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0
##	1st Qu.	: 3.5	1st Qu.	: 4.0	1st Qu.	: 2.0	1st Qu.	: 2.0
##	Median	: 11.0	Median	: 12.0	Median	: 12.0	Median	: 12.0
##	Mean	: 110.1	Mean	: 112.1	Mean	: 136.5	Mean	: 139.6
##	3rd Qu.	: 64.5	3rd Qu.	: 71.0	3rd Qu.	: 78.0	3rd Qu.	: 74.0
##	Max.	:3554.0	Max.	:3554.0	Max.	:4586.0	Max.	:4903.0
##	NA's	:16	NA's	:17	NA's	:15	NA's	:13
##	1/31/20	14:00	02/01/20	10:00	02/02/20	21:00	02/03/20	21:00
##	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0
##	1st Qu.	: 2.0	1st Qu.	: 2.0	1st Qu.	: 2.0	1st Qu.	: 2.0
##	Median	: 12.0	Median	: 12.0	Median	: 13.0	Median	: 13.0
##	Mean	: 152.7	Mean	: 179.5	Mean	: 254.3	Mean	: 292.7
##	3rd Qu.	: 83.0	3rd Qu.	: 93.5	3rd Qu.	: 107.0	3rd Qu.	: 126.0
##	Max.	:5806.0	Max.	:7153.0	Max.	:11177.0	Max.	:13522.0
##	NA's	:7	NA's	:5	NA's	:4	NA's	:3
##	02/04/20	9:40	02/04/20	22:00	02/05/20	9:00	02/05/20	23:00
##	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0
##	1st Qu.	: 2.0	1st Qu.	: 2.0	1st Qu.	: 2.0	1st Qu.	: 2.0
##	Median	: 13.5	Median	: 14.0	Median	: 14.0	Median	: 12.0
##	Mean	: 294.1	Mean	: 350.0	Mean	: 351.5	Mean	: 398.2
##	3rd Qu.	: 123.8	3rd Qu.	: 131.8	3rd Qu.	: 132.2	3rd Qu.	: 142.5
##	Max.	:13522.0	Max.	:16678.0	Max.	:16678.0	Max.	:19665.0
##	NA's	:2	NA's	:2	NA's	:2	NA's	:1
##	02/06/20	9:00	02/06/20	14:20	02/07/20	20:13	02/07/20	22:50
##	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0	Min.	: 1.0
##	1st Qu.	: 2.0	1st Qu.	: 2.0	1st Qu.	: 2.0	1st Qu.	: 2.0
##	Median	: 13.0	Median	: 16.0	Median	: 17.0	Median	: 17.0

```
## Mean : 399.2 Mean : 399.3 Mean : 437.9 Mean : 484.7
## 3rd Qu.: 145.0 3rd Qu.: 145.0 3rd Qu.: 146.5 3rd Qu.: 149.2
## Max. :19665.0 Max. :19665.0 Max. :22112.0 Max. :24953.0
## NA's :1 NA's :1
## 02/08/20 22:04 02/08/20 23:04 02/09/20 10:30 02/09/20 23:20
## Min. : 1.0 Min. : 1.0 Min. : 1.0 Min. : 1.0
## 1st Qu.: 2.0 1st Qu.: 2.0 1st Qu.: 2.0 1st Qu.: 2.0
## Median : 17.5 Median : 17.5 Median : 18.0 Median : 18.0
## Mean : 485.6 Mean : 521.5 Mean : 522.1 Mean : 563.0
## 3rd Qu.: 149.2 3rd Qu.: 153.8 3rd Qu.: 154.5 3rd Qu.: 158.2
## Max. :24953.0 Max. :27100.0 Max. :27100.0 Max. :29631.0
##
```

```
summary(ncov_recovered_df)
```

```
## Province/State Country/Region Lat Long
## Length:72 Length:72 Min. : -37.81 Min. : -123.12
## Class :character Class :character 1st Qu.: 23.80 1st Qu.: 17.12
## Mode :character Mode :character Median : 34.05 Median : 106.00
## Mean : 30.24 Mean : 64.66
## 3rd Qu.: 41.16 3rd Qu.: 117.25
## Max. : 61.92 Max. : 153.03
##
## 1/21/20 22:00 1/22/20 12:00 1/23/20 12:00 1/24/20 0:00 1/24/20 12:00
## Min. :25 Min. :28 Min. : 2.0 Min. : 1.00 Min. : 1.0
## 1st Qu.:25 1st Qu.:28 1st Qu.: 8.5 1st Qu.: 1.50 1st Qu.: 1.0
## Median :25 Median :28 Median :15.0 Median : 2.00 Median : 1.0
## Mean :25 Mean :28 Mean :15.0 Mean :11.33 Mean : 7.2
## 3rd Qu.:25 3rd Qu.:28 3rd Qu.:21.5 3rd Qu.:16.50 3rd Qu.: 2.0
## Max. :25 Max. :28 Max. :28.0 Max. :31.00 Max. :31.0
## NA's :71 NA's :71 NA's :70 NA's :69 NA's :67
## 1/25/20 0:00 1/25/20 12:00 1/25/20 22:00 1/26/20 11:00
## Min. : 1.000 Min. : 1.0 Min. : 1.000 Min. : 1.0
## 1st Qu.: 1.000 1st Qu.: 1.0 1st Qu.: 1.000 1st Qu.: 1.0
## Median : 1.000 Median : 1.5 Median : 1.500 Median : 1.5
## Mean : 6.333 Mean : 6.5 Mean : 8.167 Mean : 6.5
## 3rd Qu.: 1.750 3rd Qu.: 2.0 3rd Qu.: 2.000 3rd Qu.: 2.0
## Max. :32.000 Max. :32.0 Max. :42.000 Max. :42.0
## NA's :66 NA's :66 NA's :66 NA's :64
## 1/26/20 23:00 1/27/20 9:00 1/27/20 19:00 1/27/20 20:30
## Min. : 1.00 Min. : 1.000 Min. : 1.000 Min. : 1
## 1st Qu.: 1.00 1st Qu.: 1.000 1st Qu.: 1.000 1st Qu.: 1
## Median : 1.50 Median : 2.000 Median : 2.000 Median : 2
## Mean : 6.75 Mean : 6.556 Mean : 6.778 Mean : 7
## 3rd Qu.: 2.00 3rd Qu.: 3.000 3rd Qu.: 3.000 3rd Qu.: 3
## Max. :44.00 Max. :45.000 Max. :45.000 Max. :47
## NA's :64 NA's :63 NA's :63 NA's :63
## 1/28/20 13:00 1/28/20 18:00 1/28/20 23:00 1/29/20 13:30
## Min. : 1.00 Min. : 1.00 Min. : 1.000 Min. : 1.000
## 1st Qu.: 2.25 1st Qu.: 2.25 1st Qu.: 1.000 1st Qu.: 1.000
## Median : 3.50 Median : 3.50 Median : 3.000 Median : 2.500
## Mean : 7.90 Mean :10.70 Mean : 9.167 Mean : 8.071
## 3rd Qu.: 4.00 3rd Qu.: 4.00 3rd Qu.: 4.250 3rd Qu.: 4.000
## Max. :52.00 Max. :80.00 Max. :80.000 Max. :80.000
## NA's :62 NA's :62 NA's :60 NA's :58
```

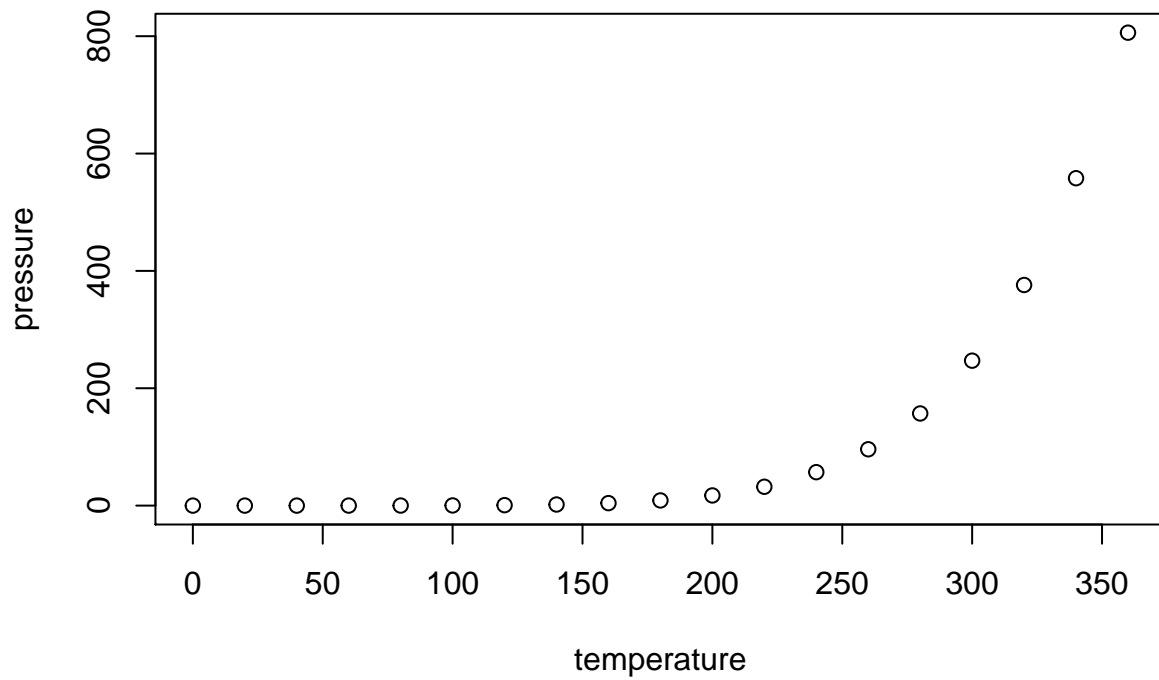
```

## 1/29/20 14:30 1/29/20 21:00 1/30/20 11:00 1/31/20 14:00
## Min. : 1.00 Min. : 1 Min. : 1.000 Min. : 1.000
## 1st Qu.: 1.00 1st Qu.: 1 1st Qu.: 1.000 1st Qu.: 1.000
## Median : 1.50 Median : 2 Median : 2.000 Median : 2.000
## Mean : 7.00 Mean : 7 Mean : 6.714 Mean : 9.565
## 3rd Qu.: 3.75 3rd Qu.: 4 3rd Qu.: 4.000 3rd Qu.: 5.000
## Max. :88.00 Max. :90 Max. :90.000 Max. :141.000
## NA's :54 NA's :53 NA's :51 NA's :49
## 02/01/20 10:00 02/02/20 21:00 02/03/20 21:00 02/04/20 9:40
## Min. : 1.00 Min. : 1.00 Min. : 0.00 Min. : 1.00
## 1st Qu.: 1.25 1st Qu.: 2.00 1st Qu.: 1.00 1st Qu.: 1.75
## Median : 3.00 Median : 4.00 Median : 3.00 Median : 4.50
## Mean : 10.92 Mean : 16.79 Mean : 17.74 Mean : 20.16
## 3rd Qu.: 7.50 3rd Qu.: 10.00 3rd Qu.: 11.00 3rd Qu.: 12.50
## Max. :168.00 Max. :295.00 Max. :386.00 Max. :396.00
## NA's :46 NA's :43 NA's :37 NA's :40
## 02/04/20 22:00 02/05/20 9:00 02/05/20 23:00 02/06/20 9:00
## Min. : 1.00 Min. : 1.00 Min. : 1.0 Min. : 1.00
## 1st Qu.: 2.00 1st Qu.: 4.00 1st Qu.: 4.0 1st Qu.: 5.00
## Median : 5.00 Median : 7.00 Median : 7.0 Median : 9.00
## Mean : 28.09 Mean : 31.03 Mean : 36.3 Mean : 41.67
## 3rd Qu.: 20.75 3rd Qu.: 23.00 3rd Qu.: 27.0 3rd Qu.: 31.00
## Max. :522.00 Max. :537.00 Max. :651.0 Max. :712.00
## NA's :40 NA's :39 NA's :39 NA's :39
## 02/06/20 14:20 02/07/20 20:13 02/07/20 22:50 02/08/20 10:24
## Min. : 1.00 Min. : 1.00 Min. : 1.00 Min. : 0.00
## 1st Qu.: 4.00 1st Qu.: 4.00 1st Qu.: 3.00 1st Qu.: 0.00
## Median : 9.00 Median : 12.00 Median : 10.00 Median : 1.00
## Mean : 39.49 Mean : 47.65 Mean : 52.50 Mean : 33.25
## 3rd Qu.: 31.00 3rd Qu.: 37.00 3rd Qu.: 33.25 3rd Qu.: 17.00
## Max. :712.00 Max. :867.00 Max. :1115.00 Max. :1218.00
## NA's :37 NA's :35 NA's :34
## 02/08/20 23:04 02/09/20 10:30 02/09/20 23:20
## Min. : 0.00 Min. : 0.00 Min. : 0.00
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.: 0.00
## Median : 1.00 Median : 1.00 Median : 1.00
## Mean : 37.51 Mean : 40.58 Mean : 46.00
## 3rd Qu.: 17.25 3rd Qu.: 18.25 3rd Qu.: 18.25
## Max. :1440.00 Max. :1480.00 Max. :1795.00
##

```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.