DataVisualization1

2023-10-02

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

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

```
library(readr)
library(knitr)
```

```
data_description <- function(df) {
  cat("First few rows of the data:\n")
  print(head(df))
  cat("\n")

  cat("Statistical summary of the data:\n")
  print(summary(df))
  cat("\n")

  cat("Structure of the data:\n")
  str(df)
  cat("\n")

  cat("Dimensions of the data:\n")
  print(dim(df))
  cat("\n")

  cat("Column names in the data:\n")
  print(colnames(df))
  cat("\n")
}
```

```
county_logalpha <- read.csv("county_logalpha.csv")
data_description(county_logalpha)
```

```
## First few rows of the data:
##   X          county.name pop_density mean_PRE sd_PRE mean_POST sd_POST
## 1 0 Autauga, Alabama, US        91.8        0      0         0       0
## 2 1 Baldwin, Alabama, US       114.6        0      0         0       0
## 3 2 Barbour, Alabama, US        31.0        0      0         0       0
## 4 3    Bibb, Alabama, US        36.8        0      0         0       0
## 5 4  Blount, Alabama, US        88.9        0      0         0       0
## 6 5 Bullock, Alabama, US        17.5        0      0         0       0
```

```
##
## Statistical summary of the data:
##        X          county.name         pop_density          mean_PRE
##  Min.   :   0.0   Length:3140        Min.   :    0.00   Min.   :  -Inf
##  1st Qu.: 784.8   Class :character   1st Qu.:   16.98   1st Qu.:0.0000
##  Median :1569.5   Mode  :character   Median :   45.20   Median :0.0000
##  Mean   :1569.5                      Mean   :  259.28   Mean   :  -Inf
##  3rd Qu.:2354.2                      3rd Qu.:  113.72   3rd Qu.:0.0000
##  Max.   :3139.0                      Max.   :69467.50   Max.   :0.6931
##                                                         NA's   :18
##      sd_PRE          mean_POST          sd_POST
##  Min.   :0.00000   Min.   :-0.05068   Min.   :0.00000
##  1st Qu.:0.00000   1st Qu.: 0.00000   1st Qu.:0.00000
##  Median :0.00000   Median : 0.00000   Median :0.00000
##  Mean   :0.01364   Mean   : 0.02180   Mean   :0.01476
##  3rd Qu.:0.00000   3rd Qu.: 0.00000   3rd Qu.:0.00000
##  Max.   :0.42105   Max.   : 0.48648   Max.   :0.38056
##  NA's   :66                           NA's   :21
##
## Structure of the data:
## 'data.frame':    3140 obs. of  7 variables:
##  $ X          : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ county.name: chr  "Autauga, Alabama, US" "Baldwin, Alabama, US" "Barbour, Alabama, US" "Bibb, Alab
##  $ pop_density: num  91.8 114.6 31 36.8 88.9 ...
##  $ mean_PRE   : num  0 0 0 0 0 ...
##  $ sd_PRE     : num  0 0 0 0 0 ...
##  $ mean_POST  : num  0 0 0 0 0 ...
##  $ sd_POST    : num  0 0 0 0 0 ...
##
## Dimensions of the data:
## [1] 3140    7
##
## Column names in the data:
## [1] "X"           "county.name" "pop_density" "mean_PRE"    "sd_PRE"
## [6] "mean_POST"   "sd_POST"
```

```
names(county_logalpha)[names(county_logalpha) == "county.name"] <- "long_county_name"
```

```
covid_county_census_data <- read.csv("covid_county_census_data.csv")
data_description(covid_county_census_data)
```

```
## First few rows of the data:
##   X fips pst045212 pst040210 pst120212 pop010210 age135212 age295212 age775212
## 1 0 1001     55514     54571       1.7     54571       6.5      26.0      13.0
## 2 1 1003    190790    182265       4.7    182265       5.9      22.6      17.7
## 3 2 1005     27201     27457      -0.9     27457       5.6      21.2      15.2
## 4 3 1007     22597     22919      -1.4     22915       5.4      21.6      13.9
## 5 4 1009     57826     57322       0.9     57322       6.2      24.0      15.9
## 6 5 1011     10474     10915      -4.0     10914       6.8      21.2      14.6
##   sex255212 rhi125212 rhi225212 rhi325212 rhi425212 rhi525212 rhi625212
## 1      51.3      78.5      18.4       0.5       1.0       0.1       1.5
## 2      51.2      87.3       9.6       0.7       0.8       0.1       1.5
## 3      46.3      50.5      47.4       0.7       0.4       0.2       0.8
## 4      46.0      76.2      22.3       0.4       0.1       0.1       0.9
## 5      50.5      96.2       1.7       0.6       0.3       0.1       1.1
```

```
## 6      45.6      27.2      69.7       0.9       0.2       0.8       1.1
##   rhi725212 rhi825212 pop715211 pop645211 pop815211 edu635211 edu685211
## 1       2.6      76.4      84.8       2.0       3.8      86.5      21.6
## 2       4.6      83.2      83.3       3.7       5.4      87.9      27.2
## 3       5.1      46.4      82.1       3.0       5.0      72.4      13.9
## 4       2.0      74.6      90.7       1.3       2.0      75.9       9.8
## 5       8.6      88.1      86.6       4.5       6.8      73.2      11.3
## 6       7.6      22.6      84.4       4.5       5.1      72.0      14.4
##   vet605211 lfe305211 hsg010211 hsg445211 hsg096211 hsg495211 hsd410211
## 1      5942      25.3     22460      77.7       7.2    137500     19998
## 2     20254      25.5    104701      76.2      23.3    175700     70757
## 3      2167      23.4     11692      66.4      10.3     91600      9589
## 4      1824      27.8      9014      83.0       6.3     87500      7225
## 5      4019      33.5     24144      79.9       4.9    111500     20954
## 6       830      28.3      4447      79.6      10.1     68000      3760
##   hsd310211 inc910211 inc110211 pvy020211 bza010211 bza110211 bza115211
## 1      2.68     25035     53899      10.9       835     10290       1.2
## 2      2.50     27217     51321      12.5      4624     51386       0.4
## 3      2.57     15899     34041      24.7       501      7572      -0.4
## 4      3.05     18462     40506      15.7       285      2980       1.8
## 5      2.70     21185     45404      13.7       690      6345      -1.4
## 6      2.78     20678     31955      26.0       113         0       0.0
##   nes010211 sbo001207 sbo315207 sbo115207 sbo215207 sbo515207 sbo415207
## 1      3062      4067      15.2       0.0       1.3         0       0.7
## 2     16097     19035       2.7       0.4       1.0         0       1.3
## 3      1513      1667       0.0       0.0       0.0         0       0.0
## 4      1193      1385      14.9       0.0       0.0         0       0.0
## 5      3544      4458       0.0       0.0       0.0         0       0.0
## 6       478       417       0.0       0.0       0.0         0       0.0
##   sbo015207 man450207 wtn220207 rtn130207 rtn131207 afn120207 bps030212
## 1      31.7         0         0    598175     12003     88157       385
## 2      27.3   1410270         0   2966490     17166    436955      1184
## 3      27.0         0         0    188337      6334         0         2
## 4       0.0         0         0    124707      5804     10757        16
## 5      23.2    341544         0    319700      5622     20941         6
## 6      38.8         0         0     43810      3995      3670         0
##   lnd110210 pop060210 fips.1    date county_name     long_county_name
## 1    594.44      91.8   1001 1/22/20     Autauga Autauga, Alabama, US
## 2   1589.78     114.6   1003 1/22/20     Baldwin Baldwin, Alabama, US
## 3    884.88      31.0   1005 1/22/20     Barbour Barbour, Alabama, US
## 4    622.58      36.8   1007 1/22/20        Bibb    Bibb, Alabama, US
## 5    644.78      88.9   1009 1/22/20      Blount  Blount, Alabama, US
## 6    622.81      17.5   1011 1/22/20     Bullock Bullock, Alabama, US
##   province_state death_count population
## 1        Alabama           0      55869
## 2        Alabama           0     223234
## 3        Alabama           0      24686
## 4        Alabama           0      22394
## 5        Alabama           0      57826
## 6        Alabama           0      10101
##
## Statistical summary of the data:
##        X               fips         pst045212         pst040210
##  Min.   :     0   Min.   : 1001   Min.   :     71   Min.   :     82
```

```
##  1st Qu.: 64370   1st Qu.:18178   1st Qu.:  11014   1st Qu.:  11118
##  Median :128740   Median :29176   Median :  25834   Median :  25890
##  Mean   :128740   Mean   :30388   Mean   :  99964   Mean   :  98319
##  3rd Qu.:193109   3rd Qu.:45080   3rd Qu.:  67355   3rd Qu.:  66898
##  Max.   :257479   Max.   :56045   Max.   :9962790   Max.   :9818600
##    pst120212          pop010210        age135212        age295212
##  Min.   :-18.1000   Min.   :     82   Min.   : 0.000   Min.   : 0.00
##  1st Qu.: -1.1000   1st Qu.:  11118   1st Qu.: 5.300   1st Qu.:20.80
##  Median : -0.1000   Median :  25890   Median : 5.900   Median :22.80
##  Mean   :  0.2127   Mean   :  98318   Mean   : 6.025   Mean   :22.85
##  3rd Qu.:  1.3000   3rd Qu.:  66898   3rd Qu.: 6.600   3rd Qu.:24.50
##  Max.   : 25.6000   Max.   :9818600   Max.   :13.300   Max.   :41.10
##    age775212       sex255212       rhi125212       rhi225212
##  Min.   : 3.60   Min.   :30.10   Min.   :10.60   Min.   : 0.000
##  1st Qu.:14.00   1st Qu.:49.50   1st Qu.:80.88   1st Qu.: 0.600
##  Median :16.40   Median :50.40   Median :92.40   Median : 2.250
##  Mean   :16.75   Mean   :49.96   Mean   :85.51   Mean   : 9.147
##  3rd Qu.:19.10   3rd Qu.:51.10   3rd Qu.:96.30   3rd Qu.:10.600
##  Max.   :49.30   Max.   :57.00   Max.   :99.20   Max.   :85.400
##    rhi325212       rhi425212        rhi525212         rhi625212
##  Min.   : 0.00   Min.   : 0.000   Min.   : 0.0000   Min.   : 0.000
##  1st Qu.: 0.30   1st Qu.: 0.400   1st Qu.: 0.0000   1st Qu.: 1.100
##  Median : 0.50   Median : 0.600   Median : 0.0000   Median : 1.400
##  Mean   : 2.14   Mean   : 1.298   Mean   : 0.1128   Mean   : 1.791
##  3rd Qu.: 1.20   3rd Qu.: 1.100   3rd Qu.: 0.1000   3rd Qu.: 2.000
##  Max.   :86.80   Max.   :43.300   Max.   :48.9000   Max.   :29.500
##    rhi725212        rhi825212       pop715211        pop645211
##  Min.   : 0.200   Min.   : 3.20   Min.   : 49.00   Min.   : 0.000
##  1st Qu.: 1.800   1st Qu.:66.20   1st Qu.: 83.70   1st Qu.: 1.200
##  Median : 3.600   Median :85.20   Median : 86.60   Median : 2.450
##  Mean   : 8.667   Mean   :77.76   Mean   : 86.14   Mean   : 4.446
##  3rd Qu.: 8.700   3rd Qu.:93.70   3rd Qu.: 89.20   3rd Qu.: 5.500
##  Max.   :95.600   Max.   :98.60   Max.   :100.00   Max.   :63.400
##    pop815211        edu635211       edu685211       vet605211
##  Min.   : 0.000   Min.   :46.30   Min.   : 4.20   Min.   :     0.0
##  1st Qu.: 2.800   1st Qu.:79.20   1st Qu.:13.30   1st Qu.:   927.8
##  Median : 4.900   Median :85.20   Median :17.10   Median :  2124.5
##  Mean   : 9.101   Mean   :83.69   Mean   :19.27   Mean   :  7074.4
##  3rd Qu.:10.100   3rd Qu.:89.10   3rd Qu.:22.80   3rd Qu.:  5801.5
##  Max.   :95.900   Max.   :98.60   Max.   :72.00   Max.   :354430.0
##    lfe305211       hsg010211         hsg445211       hsg096211
##  Min.   : 4.30   Min.   :     48   Min.   : 0.00   Min.   : 0.00
##  1st Qu.:19.20   1st Qu.:   5430   1st Qu.:69.20   1st Qu.: 6.20
##  Median :22.60   Median :  12226   Median :74.30   Median : 9.70
##  Mean   :22.86   Mean   :  42135   Mean   :73.01   Mean   :12.34
##  3rd Qu.:26.30   3rd Qu.:  30705   3rd Qu.:78.30   3rd Qu.:16.00
##  Max.   :42.50   Max.   :3449270   Max.   :93.70   Max.   :98.40
##    hsg495211        hsd410211         hsd310211       inc910211
##  Min.   :     0   Min.   :     27   Min.   :1.200   Min.   : 9412
##  1st Qu.: 81775   1st Qu.:   4242   1st Qu.:2.370   1st Qu.:19534
##  Median :108000   Median :   9821   Median :2.500   Median :22364
##  Mean   :133137   Mean   :  36546   Mean   :2.516   Mean   :23127
##  3rd Qu.:154900   3rd Qu.:  25622   3rd Qu.:2.630   3rd Qu.:25516
##  Max.   :993900   Max.   :3218520   Max.   :4.070   Max.   :61290
```

```
##    inc110211        pvy020211       bza010211          bza110211
##  Min.   : 19344   Min.   : 0.00   Min.   :     0.0   Min.   :      0
##  1st Qu.: 37793   1st Qu.:11.30   1st Qu.:   222.0   1st Qu.:   2088
##  Median : 43421   Median :15.10   Median :   536.5   Median :   6310
##  Mean   : 45304   Mean   :15.84   Mean   :  2336.1   Mean   :  35256
##  3rd Qu.: 50178   3rd Qu.:19.40   3rd Qu.:  1440.8   3rd Qu.:  19242
##  Max.   :120332   Max.   :48.40   Max.   :245261.0   Max.   :3648850
##    bza115211          nes010211          sbo001207          sbo315207
##  Min.   :-40.8000   Min.   :     0.0   Min.   :      0.0   Min.   : 0.000
##  1st Qu.: -2.1000   1st Qu.:   750.8   1st Qu.:    904.8   1st Qu.: 0.000
##  Median :  0.2000   Median :  1657.0   Median :   2159.0   Median : 0.000
##  Mean   :  0.5151   Mean   :  7162.3   Mean   :   8788.7   Mean   : 2.393
##  3rd Qu.:  2.6000   3rd Qu.:  4259.2   3rd Qu.:   5643.0   3rd Qu.: 0.000
##  Max.   :288.2000   Max.   :904398.0   Max.   :1046940.0   Max.   :66.700
##    sbo115207          sbo215207         sbo515207           sbo415207
##  Min.   : 0.0000   Min.   : 0.000   Min.   : 0.00000   Min.   : 0.000
##  1st Qu.: 0.0000   1st Qu.: 0.000   1st Qu.: 0.00000   1st Qu.: 0.000
##  Median : 0.0000   Median : 0.000   Median : 0.00000   Median : 0.000
##  Mean   : 0.5752   Mean   : 0.801   Mean   : 0.01599   Mean   : 1.692
##  3rd Qu.: 0.0000   3rd Qu.: 0.000   3rd Qu.: 0.00000   3rd Qu.: 0.000
##  Max.   :71.8000   Max.   :56.600   Max.   :10.50000   Max.   :78.000
##    sbo015207        man450207           wtn220207           rtn130207
##  Min.   : 0.00   Min.   :        0   Min.   :        0   Min.   :        0
##  1st Qu.: 0.00   1st Qu.:        0   1st Qu.:        0   1st Qu.:    76040
##  Median :23.20   Median :        0   Median :    44379   Median :   247782
##  Mean   :17.96   Mean   :  1421028   Mean   :  1211975   Mean   :  1246527
##  3rd Qu.:27.80   3rd Qu.:   896601   3rd Qu.:   249462   3rd Qu.:   767070
##  Max.   :56.20   Max.   :169275000   Max.   :205479000   Max.   :119112000
##    rtn131207        afn120207          bps030212        lnd110210
##  Min.   :    0   Min.   :       0   Min.   :    0   Min.   :     2.0
##  1st Qu.: 6859   1st Qu.:    6068   1st Qu.:    6   1st Qu.:   430.7
##  Median : 9729   Median :   25158   Median :   30   Median :   615.6
##  Mean   :10237   Mean   :  193079   Mean   :  264   Mean   :  1118.7
##  3rd Qu.:12933   3rd Qu.:   93528   3rd Qu.:  128   3rd Qu.:   923.6
##  Max.   :80800   Max.   :24857800   Max.   :27154   Max.   :145505.0
##    pop060210            fips.1          date            county_name
##  Min.   :    0.00   Min.   : 1001   Length:257480      Length:257480
##  1st Qu.:   16.98   1st Qu.:18178   Class :character   Class :character
##  Median :   45.20   Median :29176   Mode  :character   Mode  :character
##  Mean   :  259.28   Mean   :30388
##  3rd Qu.:  113.72   3rd Qu.:45080
##  Max.   :69467.50   Max.   :56045
##  long_county_name   province_state      death_count          population
##  Length:257480      Length:257480      Min.   :   0.000   Min.   :      86
##  Class :character   Class :character   1st Qu.:   0.000   1st Qu.:   10906
##  Mode  :character   Mode  :character   Median :   0.000   Median :   25764
##                                        Mean   :   0.719   Mean   :  105857
##                                        3rd Qu.:   0.000   3rd Qu.:   68104
##                                        Max.   :7349.000   Max.   :10039107
##
## Structure of the data:
## 'data.frame':    257480 obs. of  60 variables:
##  $ X               : int  0 1 2 3 4 5 6 7 8 9 ...
##  $ fips            : int  1001 1003 1005 1007 1009 1011 1013 1015 1017 1019 ...
```

```
##  $ pst045212       : num  55514 190790 27201 22597 57826 ...
##  $ pst040210       : num  54571 182265 27457 22919 57322 ...
##  $ pst120212       : num  1.7 4.7 -0.9 -1.4 0.9 -4 -3.1 -1.1 -0.4 0.1 ...
##  $ pop010210       : num  54571 182265 27457 22915 57322 ...
##  $ age135212       : num  6.5 5.9 5.6 5.4 6.2 6.8 6.1 6.1 5.8 4.8 ...
##  $ age295212       : num  26 22.6 21.2 21.6 24 21.2 23.1 22.7 22 20.8 ...
##  $ age775212       : num  13 17.7 15.2 13.9 15.9 14.6 17.4 15.2 17.6 19.3 ...
##  $ sex255212       : num  51.3 51.2 46.3 46 50.5 45.6 53.2 51.8 52.3 50.3 ...
##  $ rhi125212       : num  78.5 87.3 50.5 76.2 96.2 27.2 54.4 76 58.9 93.2 ...
##  $ rhi225212       : num  18.4 9.6 47.4 22.3 1.7 69.7 43.5 21 39.1 4.6 ...
##  $ rhi325212       : num  0.5 0.7 0.7 0.4 0.6 0.9 0.3 0.5 0.2 0.5 ...
##  $ rhi425212       : num  1 0.8 0.4 0.1 0.3 0.2 1 0.8 0.6 0.3 ...
##  $ rhi525212       : num  0.1 0.1 0.2 0.1 0.1 0.8 0 0.1 0.1 0 ...
##  $ rhi625212       : num  1.5 1.5 0.8 0.9 1.1 1.1 0.7 1.6 1.1 1.4 ...
##  $ rhi725212       : num  2.6 4.6 5.1 2 8.6 7.6 1.2 3.5 1.9 1.4 ...
##  $ rhi825212       : num  76.4 83.2 46.4 74.6 88.1 22.6 53.7 73.2 57.6 91.9 ...
##  $ pop715211       : num  84.8 83.3 82.1 90.7 86.6 84.4 93 83.5 84.8 89.3 ...
##  $ pop645211       : num  2 3.7 3 1.3 4.5 4.5 1 2.3 1.1 0.8 ...
##  $ pop815211       : num  3.8 5.4 5 2 6.8 5.1 1.6 4.5 1.8 1.3 ...
##  $ edu635211       : num  86.5 87.9 72.4 75.9 73.2 72 74.7 77.9 74.2 73.5 ...
##  $ edu685211       : num  21.6 27.2 13.9 9.8 11.3 14.4 12.4 15.8 10.5 11.3 ...
##  $ vet605211       : num  5942 20254 2167 1824 4019 ...
##  $ lfe305211       : num  25.3 25.5 23.4 27.8 33.5 28.3 23.9 22.2 24.1 25.8 ...
##  $ hsg010211       : num  22460 104701 11692 9014 24144 ...
##  $ hsg445211       : num  77.7 76.2 66.4 83 79.9 79.6 71.1 70.1 71.4 75.8 ...
##  $ hsg096211       : num  7.2 23.3 10.3 6.3 4.9 10.1 13.8 13.8 10.6 4.5 ...
##  $ hsg495211       : num  137500 175700 91600 87500 111500 ...
##  $ hsd410211       : num  19998 70757 9589 7225 20954 ...
##  $ hsd310211       : num  2.68 2.5 2.57 3.05 2.7 2.78 2.55 2.5 2.5 2.2 ...
##  $ inc910211       : num  25035 27217 15899 18462 21185 ...
##  $ inc110211       : num  53899 51321 34041 40506 45404 ...
##  $ pvy020211       : num  10.9 12.5 24.7 15.7 13.7 26 25.7 20.4 21 20.8 ...
##  $ bza010211       : num  835 4624 501 285 690 ...
##  $ bza110211       : num  10290 51386 7572 2980 6345 ...
##  $ bza115211       : num  1.2 0.4 -0.4 1.8 -1.4 0 -3.3 -2.3 4.6 3.6 ...
##  $ nes010211       : num  3062 16097 1513 1193 3544 ...
##  $ sbo001207       : num  4067 19035 1667 1385 4458 ...
##  $ sbo315207       : num  15.2 2.7 0 14.9 0 0 0 7.2 0 0 ...
##  $ sbo115207       : num  0 0.4 0 0 0 0 0 0 0 0 ...
##  $ sbo215207       : num  1.3 1 0 0 0 0 3.3 1.6 0 0 ...
##  $ sbo515207       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sbo415207       : num  0.7 1.3 0 0 0 0 0 0.5 0 0 ...
##  $ sbo015207       : num  31.7 27.3 27 0 23.2 38.8 0 24.7 29.3 14.5 ...
##  $ man450207       : num  0 1410270 0 0 341544 ...
##  $ wtn220207       : num  0 0 0 0 0 ...
##  $ rtn130207       : num  598175 2966490 188337 124707 319700 ...
##  $ rtn131207       : num  12003 17166 6334 5804 5622 ...
##  $ afn120207       : num  88157 436955 0 10757 20941 ...
##  $ bps030212       : num  385 1184 2 16 6 ...
##  $ lnd110210       : num  594 1590 885 623 645 ...
##  $ pop060210       : num  91.8 114.6 31 36.8 88.9 ...
##  $ fips.1          : int  1001 1003 1005 1007 1009 1011 1013 1015 1017 1019 ...
##  $ date            : chr  "1/22/20" "1/22/20" "1/22/20" "1/22/20" ...
##  $ county_name     : chr  "Autauga" "Baldwin" "Barbour" "Bibb" ...
```

```
##  $ long_county_name: chr  "Autauga, Alabama, US" "Baldwin, Alabama, US" "Barbour, Alabama, US" "Bibb
##  $ province_state  : chr  "Alabama" "Alabama" "Alabama" "Alabama" ...
##  $ death_count     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ population      : int  55869 223234 24686 22394 57826 10101 19448 113605 33254 26196 ...
##
## Dimensions of the data:
## [1] 257480     60
##
## Column names in the data:
##  [1] "X"                "fips"             "pst045212"        "pst040210"
##  [5] "pst120212"        "pop010210"        "age135212"        "age295212"
##  [9] "age775212"        "sex255212"        "rhi125212"        "rhi225212"
## [13] "rhi325212"        "rhi425212"        "rhi525212"        "rhi625212"
## [17] "rhi725212"        "rhi825212"        "pop715211"        "pop645211"
## [21] "pop815211"        "edu635211"        "edu685211"        "vet605211"
## [25] "lfe305211"        "hsg010211"        "hsg445211"        "hsg096211"
## [29] "hsg495211"        "hsd410211"        "hsd310211"        "inc910211"
## [33] "inc110211"        "pvy020211"        "bza010211"        "bza110211"
## [37] "bza115211"        "nes010211"        "sbo001207"        "sbo315207"
## [41] "sbo115207"        "sbo215207"        "sbo515207"        "sbo415207"
## [45] "sbo015207"        "man450207"        "wtn220207"        "rtn130207"
## [49] "rtn131207"        "afn120207"        "bps030212"        "lnd110210"
## [53] "pop060210"        "fips.1"           "date"             "county_name"
## [57] "long_county_name" "province_state"   "death_count"      "population"
```

```r
covid_data <- merge(county_logalpha, covid_county_census_data, by = "long_county_name")
data_description(covid_data)
```

```
## First few rows of the data:
##                long_county_name  X.x pop_density mean_PRE sd_PRE mean_POST
## 1 Abbeville, South Carolina, US 2315        51.8        0      0         0
## 2 Abbeville, South Carolina, US 2315        51.8        0      0         0
## 3 Abbeville, South Carolina, US 2315        51.8        0      0         0
## 4 Abbeville, South Carolina, US 2315        51.8        0      0         0
## 5 Abbeville, South Carolina, US 2315        51.8        0      0         0
## 6 Abbeville, South Carolina, US 2315        51.8        0      0         0
##   sd_POST    X.y  fips pst045212 pst040210 pst120212 pop010210 age135212
## 1       0 171875 45001     25101     25421      -1.3     25417       5.6
## 2       0  83955 45001     25101     25421      -1.3     25417       5.6
## 3       0 253515 45001     25101     25421      -1.3     25417       5.6
## 4       0 237815 45001     25101     25421      -1.3     25417       5.6
## 5       0 121635 45001     25101     25421      -1.3     25417       5.6
## 6       0  90235 45001     25101     25421      -1.3     25417       5.6
##   age295212 age775212 sex255212 rhi125212 rhi225212 rhi325212 rhi425212
## 1      22.2      17.9      51.5      69.9      28.2       0.3       0.4
## 2      22.2      17.9      51.5      69.9      28.2       0.3       0.4
## 3      22.2      17.9      51.5      69.9      28.2       0.3       0.4
## 4      22.2      17.9      51.5      69.9      28.2       0.3       0.4
## 5      22.2      17.9      51.5      69.9      28.2       0.3       0.4
## 6      22.2      17.9      51.5      69.9      28.2       0.3       0.4
##   rhi525212 rhi625212 rhi725212 rhi825212 pop715211 pop645211 pop815211
## 1         0       1.2       1.2      68.9      90.1       1.5       2.3
## 2         0       1.2       1.2      68.9      90.1       1.5       2.3
## 3         0       1.2       1.2      68.9      90.1       1.5       2.3
## 4         0       1.2       1.2      68.9      90.1       1.5       2.3
```

```
## 5         0       1.2       1.2      68.9      90.1       1.5       2.3
## 6         0       1.2       1.2      68.9      90.1       1.5       2.3
##   edu635211 edu685211 vet605211 lfe305211 hsg010211 hsg445211 hsg096211
## 1      77.1      13.9      2019        26     12072      77.9       6.9
## 2      77.1      13.9      2019        26     12072      77.9       6.9
## 3      77.1      13.9      2019        26     12072      77.9       6.9
## 4      77.1      13.9      2019        26     12072      77.9       6.9
## 5      77.1      13.9      2019        26     12072      77.9       6.9
## 6      77.1      13.9      2019        26     12072      77.9       6.9
##   hsg495211 hsd410211 hsd310211 inc910211 inc110211 pvy020211 bza010211
## 1     89100      9811      2.53     17424     34670      19.6       334
## 2     89100      9811      2.53     17424     34670      19.6       334
## 3     89100      9811      2.53     17424     34670      19.6       334
## 4     89100      9811      2.53     17424     34670      19.6       334
## 5     89100      9811      2.53     17424     34670      19.6       334
## 6     89100      9811      2.53     17424     34670      19.6       334
##   bza110211 bza115211 nes010211 sbo001207 sbo315207 sbo115207 sbo215207
## 1      4119      -4.1      1472      1385      19.1         0         0
## 2      4119      -4.1      1472      1385      19.1         0         0
## 3      4119      -4.1      1472      1385      19.1         0         0
## 4      4119      -4.1      1472      1385      19.1         0         0
## 5      4119      -4.1      1472      1385      19.1         0         0
## 6      4119      -4.1      1472      1385      19.1         0         0
##   sbo515207 sbo415207 sbo015207 man450207 wtn220207 rtn130207 rtn131207
## 1         0         0      33.4    657498         0     71936      2841
## 2         0         0      33.4    657498         0     71936      2841
## 3         0         0      33.4    657498         0     71936      2841
## 4         0         0      33.4    657498         0     71936      2841
## 5         0         0      33.4    657498         0     71936      2841
## 6         0         0      33.4    657498         0     71936      2841
##   afn120207 bps030212 lnd110210 pop060210 fips.1    date county_name
## 1     10963        23    490.48      51.8  45001 3/17/20   Abbeville
## 2     10963        23    490.48      51.8  45001 2/17/20   Abbeville
## 3     10963        23    490.48      51.8  45001 4/12/20   Abbeville
## 4     10963        23    490.48      51.8  45001  4/7/20   Abbeville
## 5     10963        23    490.48      51.8  45001  3/1/20   Abbeville
## 6     10963        23    490.48      51.8  45001 2/19/20   Abbeville
##   province_state death_count population
## 1 South Carolina           0      24527
## 2 South Carolina           0      24527
## 3 South Carolina           0      24527
## 4 South Carolina           0      24527
## 5 South Carolina           0      24527
## 6 South Carolina           0      24527
##
## Statistical summary of the data:
##  long_county_name        X.x          pop_density          mean_PRE
##  Length:257480      Min.   :   0.0   Min.   :    0.00   Min.   :  -Inf
##  Class :character   1st Qu.: 784.8   1st Qu.:   16.98   1st Qu.:0.0000
##  Mode  :character   Median :1569.5   Median :   45.20   Median :0.0000
##                     Mean   :1569.5   Mean   :  259.28   Mean   :  -Inf
##                     3rd Qu.:2354.2   3rd Qu.:  113.72   3rd Qu.:0.0000
##                     Max.   :3139.0   Max.   :69467.50   Max.   :0.6931
##                                                         NA's   :1476
```

```
##      sd_PRE        mean_POST           sd_POST            X.y
##  Min.   :0.000   Min.   :-0.05068   Min.   :0.0000   Min.   :     0
##  1st Qu.:0.000   1st Qu.: 0.00000   1st Qu.:0.0000   1st Qu.: 64370
##  Median :0.000   Median : 0.00000   Median :0.0000   Median :128740
##  Mean   :0.014   Mean   : 0.02180   Mean   :0.0148   Mean   :128740
##  3rd Qu.:0.000   3rd Qu.: 0.00000   3rd Qu.:0.0000   3rd Qu.:193109
##  Max.   :0.421   Max.   : 0.48648   Max.   :0.3806   Max.   :257479
##  NA's   :5412                       NA's   :1722
##       fips         pst045212         pst040210         pst120212
##  Min.   : 1001   Min.   :     71   Min.   :     82   Min.   :-18.1000
##  1st Qu.:18178   1st Qu.:  11014   1st Qu.:  11118   1st Qu.: -1.1000
##  Median :29176   Median :  25834   Median :  25890   Median : -0.1000
##  Mean   :30388   Mean   :  99964   Mean   :  98319   Mean   :  0.2127
##  3rd Qu.:45080   3rd Qu.:  67355   3rd Qu.:  66898   3rd Qu.:  1.3000
##  Max.   :56045   Max.   :9962790   Max.   :9818600   Max.   : 25.6000
##
##    pop010210         age135212        age295212       age775212
##  Min.   :     82   Min.   : 0.000   Min.   : 0.00   Min.   : 3.60
##  1st Qu.:  11118   1st Qu.: 5.300   1st Qu.:20.80   1st Qu.:14.00
##  Median :  25890   Median : 5.900   Median :22.80   Median :16.40
##  Mean   :  98318   Mean   : 6.025   Mean   :22.85   Mean   :16.75
##  3rd Qu.:  66898   3rd Qu.: 6.600   3rd Qu.:24.50   3rd Qu.:19.10
##  Max.   :9818600   Max.   :13.300   Max.   :41.10   Max.   :49.30
##
##    sex255212       rhi125212       rhi225212        rhi325212
##  Min.   :30.10   Min.   :10.60   Min.   : 0.000   Min.   : 0.00
##  1st Qu.:49.50   1st Qu.:80.88   1st Qu.: 0.600   1st Qu.: 0.30
##  Median :50.40   Median :92.40   Median : 2.250   Median : 0.50
##  Mean   :49.96   Mean   :85.51   Mean   : 9.147   Mean   : 2.14
##  3rd Qu.:51.10   3rd Qu.:96.30   3rd Qu.:10.600   3rd Qu.: 1.20
##  Max.   :57.00   Max.   :99.20   Max.   :85.400   Max.   :86.80
##
##    rhi425212        rhi525212         rhi625212        rhi725212
##  Min.   : 0.000   Min.   : 0.0000   Min.   : 0.000   Min.   : 0.200
##  1st Qu.: 0.400   1st Qu.: 0.0000   1st Qu.: 1.100   1st Qu.: 1.800
##  Median : 0.600   Median : 0.0000   Median : 1.400   Median : 3.600
##  Mean   : 1.298   Mean   : 0.1128   Mean   : 1.791   Mean   : 8.667
##  3rd Qu.: 1.100   3rd Qu.: 0.1000   3rd Qu.: 2.000   3rd Qu.: 8.700
##  Max.   :43.300   Max.   :48.9000   Max.   :29.500   Max.   :95.600
##
##    rhi825212       pop715211        pop645211        pop815211
##  Min.   : 3.20   Min.   : 49.00   Min.   : 0.000   Min.   : 0.000
##  1st Qu.:66.20   1st Qu.: 83.70   1st Qu.: 1.200   1st Qu.: 2.800
##  Median :85.20   Median : 86.60   Median : 2.450   Median : 4.900
##  Mean   :77.76   Mean   : 86.14   Mean   : 4.446   Mean   : 9.101
##  3rd Qu.:93.70   3rd Qu.: 89.20   3rd Qu.: 5.500   3rd Qu.:10.100
##  Max.   :98.60   Max.   :100.00   Max.   :63.400   Max.   :95.900
##
##    edu635211       edu685211       vet605211          lfe305211
##  Min.   :46.30   Min.   : 4.20   Min.   :     0.0   Min.   : 4.30
##  1st Qu.:79.20   1st Qu.:13.30   1st Qu.:   927.8   1st Qu.:19.20
##  Median :85.20   Median :17.10   Median :  2124.5   Median :22.60
##  Mean   :83.69   Mean   :19.27   Mean   :  7074.4   Mean   :22.86
##  3rd Qu.:89.10   3rd Qu.:22.80   3rd Qu.:  5801.5   3rd Qu.:26.30
```

```
##  Max.   :98.60   Max.   :72.00   Max.   :354430.0   Max.   :42.50
##
##    hsg010211         hsg445211       hsg096211       hsg495211
##  Min.   :     48   Min.   : 0.00   Min.   : 0.00   Min.   :     0
##  1st Qu.:   5430   1st Qu.:69.20   1st Qu.: 6.20   1st Qu.: 81775
##  Median :  12226   Median :74.30   Median : 9.70   Median :108000
##  Mean   :  42135   Mean   :73.01   Mean   :12.34   Mean   :133137
##  3rd Qu.:  30705   3rd Qu.:78.30   3rd Qu.:16.00   3rd Qu.:154900
##  Max.   :3449270   Max.   :93.70   Max.   :98.40   Max.   :993900
##
##    hsd410211         hsd310211       inc910211       inc110211
##  Min.   :     27   Min.   :1.200   Min.   : 9412   Min.   : 19344
##  1st Qu.:   4242   1st Qu.:2.370   1st Qu.:19534   1st Qu.: 37793
##  Median :   9821   Median :2.500   Median :22364   Median : 43421
##  Mean   :  36546   Mean   :2.516   Mean   :23127   Mean   : 45304
##  3rd Qu.:  25622   3rd Qu.:2.630   3rd Qu.:25516   3rd Qu.: 50178
##  Max.   :3218520   Max.   :4.070   Max.   :61290   Max.   :120332
##
##    pvy020211       bza010211          bza110211         bza115211
##  Min.   : 0.00   Min.   :     0.0   Min.   :      0   Min.   :-40.8000
##  1st Qu.:11.30   1st Qu.:   222.0   1st Qu.:   2088   1st Qu.: -2.1000
##  Median :15.10   Median :   536.5   Median :   6310   Median :  0.2000
##  Mean   :15.84   Mean   :  2336.1   Mean   :  35256   Mean   :  0.5151
##  3rd Qu.:19.40   3rd Qu.:  1440.8   3rd Qu.:  19242   3rd Qu.:  2.6000
##  Max.   :48.40   Max.   :245261.0   Max.   :3648850   Max.   :288.2000
##
##    nes010211          sbo001207          sbo315207        sbo115207
##  Min.   :     0.0   Min.   :      0.0   Min.   : 0.000   Min.   : 0.0000
##  1st Qu.:   750.8   1st Qu.:    904.8   1st Qu.: 0.000   1st Qu.: 0.0000
##  Median :  1657.0   Median :   2159.0   Median : 0.000   Median : 0.0000
##  Mean   :  7162.3   Mean   :   8788.7   Mean   : 2.393   Mean   : 0.5752
##  3rd Qu.:  4259.2   3rd Qu.:   5643.0   3rd Qu.: 0.000   3rd Qu.: 0.0000
##  Max.   :904398.0   Max.   :1046940.0   Max.   :66.700   Max.   :71.8000
##
##    sbo215207        sbo515207          sbo415207        sbo015207
##  Min.   : 0.000   Min.   : 0.00000   Min.   : 0.000   Min.   : 0.00
##  1st Qu.: 0.000   1st Qu.: 0.00000   1st Qu.: 0.000   1st Qu.: 0.00
##  Median : 0.000   Median : 0.00000   Median : 0.000   Median :23.20
##  Mean   : 0.801   Mean   : 0.01599   Mean   : 1.692   Mean   :17.96
##  3rd Qu.: 0.000   3rd Qu.: 0.00000   3rd Qu.: 0.000   3rd Qu.:27.80
##  Max.   :56.600   Max.   :10.50000   Max.   :78.000   Max.   :56.20
##
##    man450207           wtn220207           rtn130207           rtn131207
##  Min.   :        0   Min.   :        0   Min.   :        0   Min.   :    0
##  1st Qu.:        0   1st Qu.:        0   1st Qu.:    76040   1st Qu.: 6859
##  Median :        0   Median :    44379   Median :   247782   Median : 9729
##  Mean   :  1421028   Mean   :  1211975   Mean   :  1246527   Mean   :10237
##  3rd Qu.:   896601   3rd Qu.:   249462   3rd Qu.:   767070   3rd Qu.:12933
##  Max.   :169275000   Max.   :205479000   Max.   :119112000   Max.   :80800
##
##    afn120207           bps030212       lnd110210          pop060210
##  Min.   :       0   Min.   :    0   Min.   :     2.0   Min.   :    0.00
##  1st Qu.:    6068   1st Qu.:    6   1st Qu.:   430.7   1st Qu.:   16.98
##  Median :   25158   Median :   30   Median :   615.6   Median :   45.20
```

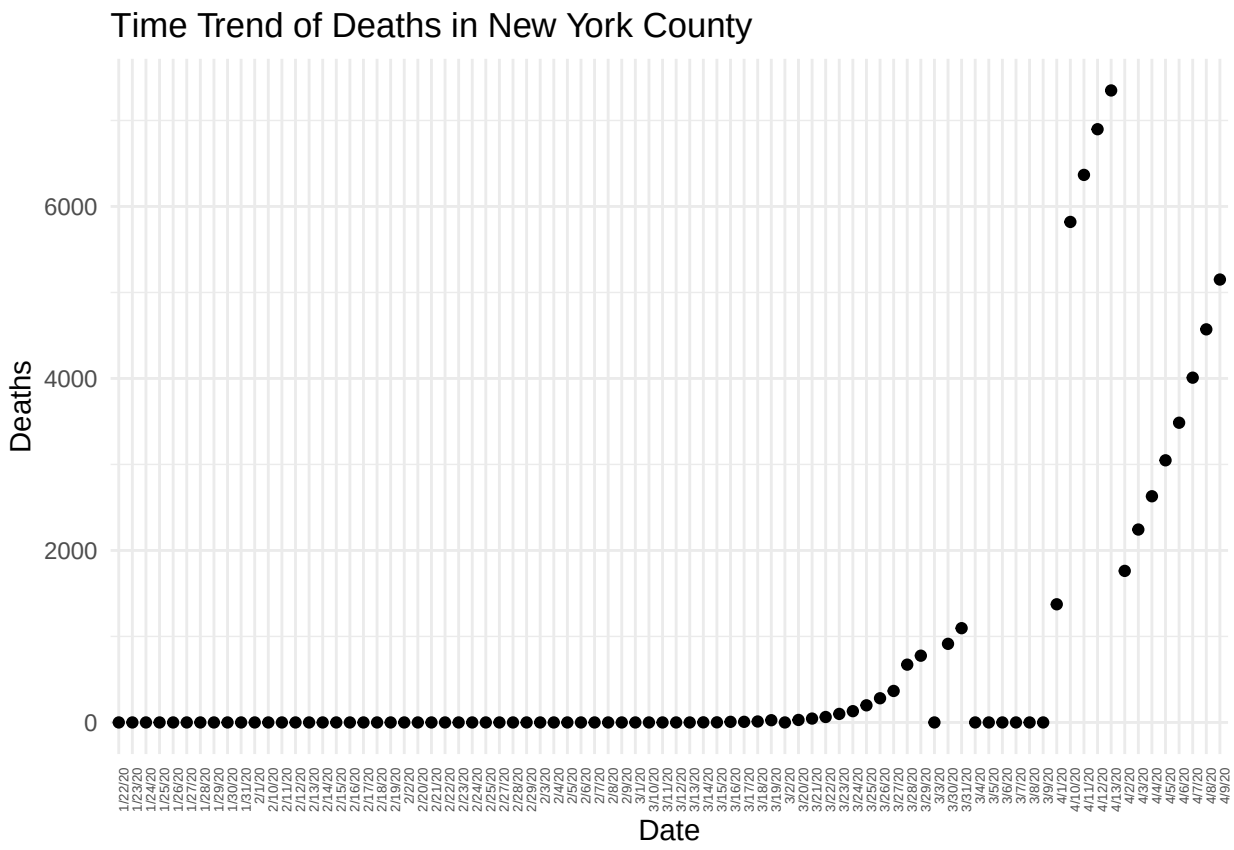
```
##  Mean   :  193079   Mean   :  264   Mean   :  1118.7   Mean   :  259.28
##  3rd Qu.:   93528   3rd Qu.:  128   3rd Qu.:   923.6   3rd Qu.:  113.72
##  Max.   :24857800   Max.   :27154   Max.   :145505.0   Max.   :69467.50
##
##      fips.1          date           county_name        province_state
##  Min.   : 1001   Length:257480      Length:257480      Length:257480
##  1st Qu.:18178   Class :character   Class :character   Class :character
##  Median :29176   Mode  :character   Mode  :character   Mode  :character
##  Mean   :30388
##  3rd Qu.:45080
##  Max.   :56045
##
##   death_count         population
##  Min.   :   0.000   Min.   :      86
##  1st Qu.:   0.000   1st Qu.:   10906
##  Median :   0.000   Median :   25764
##  Mean   :   0.719   Mean   :  105857
##  3rd Qu.:   0.000   3rd Qu.:   68104
##  Max.   :7349.000   Max.   :10039107
##
##
## Structure of the data:
## 'data.frame':    257480 obs. of  66 variables:
##  $ long_county_name: chr  "Abbeville, South Carolina, US" "Abbeville, South Carolina, US" "Abbeville
##  $ X.x             : int  2315 2315 2315 2315 2315 2315 2315 2315 2315 2315 ...
##  $ pop_density     : num  51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8 ...
##  $ mean_PRE        : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sd_PRE          : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ mean_POST       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sd_POST         : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ X.y             : int  171875 83955 253515 237815 121635 90235 140475 93375 231535 77675 ...
##  $ fips            : int  45001 45001 45001 45001 45001 45001 45001 45001 45001 45001 ...
##  $ pst045212       : num  25101 25101 25101 25101 25101 ...
##  $ pst040210       : num  25421 25421 25421 25421 25421 ...
##  $ pst120212       : num  -1.3 -1.3 -1.3 -1.3 -1.3 -1.3 -1.3 -1.3 -1.3 -1.3 ...
##  $ pop010210       : num  25417 25417 25417 25417 25417 ...
##  $ age135212       : num  5.6 5.6 5.6 5.6 5.6 5.6 5.6 5.6 5.6 5.6 ...
##  $ age295212       : num  22.2 22.2 22.2 22.2 22.2 22.2 22.2 22.2 22.2 22.2 ...
##  $ age775212       : num  17.9 17.9 17.9 17.9 17.9 17.9 17.9 17.9 17.9 17.9 ...
##  $ sex255212       : num  51.5 51.5 51.5 51.5 51.5 51.5 51.5 51.5 51.5 51.5 ...
##  $ rhi125212       : num  69.9 69.9 69.9 69.9 69.9 69.9 69.9 69.9 69.9 69.9 ...
##  $ rhi225212       : num  28.2 28.2 28.2 28.2 28.2 28.2 28.2 28.2 28.2 28.2 ...
##  $ rhi325212       : num  0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 0.3 ...
##  $ rhi425212       : num  0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 ...
##  $ rhi525212       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ rhi625212       : num  1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 ...
##  $ rhi725212       : num  1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 1.2 ...
##  $ rhi825212       : num  68.9 68.9 68.9 68.9 68.9 68.9 68.9 68.9 68.9 68.9 ...
##  $ pop715211       : num  90.1 90.1 90.1 90.1 90.1 90.1 90.1 90.1 90.1 90.1 ...
##  $ pop645211       : num  1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 1.5 ...
##  $ pop815211       : num  2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 2.3 ...
##  $ edu635211       : num  77.1 77.1 77.1 77.1 77.1 77.1 77.1 77.1 77.1 77.1 ...
##  $ edu685211       : num  13.9 13.9 13.9 13.9 13.9 13.9 13.9 13.9 13.9 13.9 ...
##  $ vet605211       : num  2019 2019 2019 2019 2019 ...
```

```
##  $ lfe305211       : num  26 26 26 26 26 26 26 26 26 26 ...
##  $ hsg010211       : num  12072 12072 12072 12072 12072 ...
##  $ hsg445211       : num  77.9 77.9 77.9 77.9 77.9 77.9 77.9 77.9 77.9 77.9 ...
##  $ hsg096211       : num  6.9 6.9 6.9 6.9 6.9 6.9 6.9 6.9 6.9 6.9 ...
##  $ hsg495211       : num  89100 89100 89100 89100 89100 89100 89100 89100 89100 89100 ...
##  $ hsd410211       : num  9811 9811 9811 9811 9811 ...
##  $ hsd310211       : num  2.53 2.53 2.53 2.53 2.53 2.53 2.53 2.53 2.53 2.53 ...
##  $ inc910211       : num  17424 17424 17424 17424 17424 ...
##  $ inc110211       : num  34670 34670 34670 34670 34670 ...
##  $ pvy020211       : num  19.6 19.6 19.6 19.6 19.6 19.6 19.6 19.6 19.6 19.6 ...
##  $ bza010211       : num  334 334 334 334 334 334 334 334 334 334 ...
##  $ bza110211       : num  4119 4119 4119 4119 4119 ...
##  $ bza115211       : num  -4.1 -4.1 -4.1 -4.1 -4.1 -4.1 -4.1 -4.1 -4.1 -4.1 ...
##  $ nes010211       : num  1472 1472 1472 1472 1472 ...
##  $ sbo001207       : num  1385 1385 1385 1385 1385 ...
##  $ sbo315207       : num  19.1 19.1 19.1 19.1 19.1 19.1 19.1 19.1 19.1 19.1 ...
##  $ sbo115207       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sbo215207       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sbo515207       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sbo415207       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ sbo015207       : num  33.4 33.4 33.4 33.4 33.4 33.4 33.4 33.4 33.4 33.4 ...
##  $ man450207       : num  657498 657498 657498 657498 657498 ...
##  $ wtn220207       : num  0 0 0 0 0 0 0 0 0 0 ...
##  $ rtn130207       : num  71936 71936 71936 71936 71936 ...
##  $ rtn131207       : num  2841 2841 2841 2841 2841 ...
##  $ afn120207       : num  10963 10963 10963 10963 10963 ...
##  $ bps030212       : num  23 23 23 23 23 23 23 23 23 23 ...
##  $ lnd110210       : num  490 490 490 490 490 ...
##  $ pop060210       : num  51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8 51.8 ...
##  $ fips.1          : int  45001 45001 45001 45001 45001 45001 45001 45001 45001 45001 ...
##  $ date            : chr  "3/17/20" "2/17/20" "4/12/20" "4/7/20" ...
##  $ county_name     : chr  "Abbeville" "Abbeville" "Abbeville" "Abbeville" ...
##  $ province_state  : chr  "South Carolina" "South Carolina" "South Carolina" "South Carolina" ...
##  $ death_count     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ population      : int  24527 24527 24527 24527 24527 24527 24527 24527 24527 24527 ...
##
## Dimensions of the data:
## [1] 257480     66
##
## Column names in the data:
##  [1] "long_county_name" "X.x"              "pop_density"      "mean_PRE"
##  [5] "sd_PRE"           "mean_POST"        "sd_POST"          "X.y"
##  [9] "fips"             "pst045212"        "pst040210"        "pst120212"
## [13] "pop010210"        "age135212"        "age295212"        "age775212"
## [17] "sex255212"        "rhi125212"        "rhi225212"        "rhi325212"
## [21] "rhi425212"        "rhi525212"        "rhi625212"        "rhi725212"
## [25] "rhi825212"        "pop715211"        "pop645211"        "pop815211"
## [29] "edu635211"        "edu685211"        "vet605211"        "lfe305211"
## [33] "hsg010211"        "hsg445211"        "hsg096211"        "hsg495211"
## [37] "hsd410211"        "hsd310211"        "inc910211"        "inc110211"
## [41] "pvy020211"        "bza010211"        "bza110211"        "bza115211"
## [45] "nes010211"        "sbo001207"        "sbo315207"        "sbo115207"
## [49] "sbo215207"        "sbo515207"        "sbo415207"        "sbo015207"
## [53] "man450207"        "wtn220207"        "rtn130207"        "rtn131207"
```

```
## [57] "afn120207"        "bps030212"        "lnd110210"        "pop060210"
## [61] "fips.1"           "date"             "county_name"      "province_state"
## [65] "death_count"      "population"
```

```
ny <- covid_data[covid_data$county_name == 'New York', ]
```

```
library(ggplot2)
ggplot(data = ny, aes(x = date, y = death_count)) +
  geom_line() +
  geom_point() +
  theme_minimal() +
  labs(title = "Time Trend of Deaths in New York County",
       x = "Date",
       y = "Deaths") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1, size = 5))
```




```
df <- read.csv('case-hosp-death.csv', header=TRUE)
data_description(df)
```

```
## First few rows of the data:
##   DATE_OF_INTEREST CASE_COUNT HOSPITALIZED_COUNT DEATH_COUNT
## 1       02/29/2020          1                 11           0
## 2       03/01/2020          0                  4           0
## 3       03/02/2020          0                 21           0
## 4       03/03/2020          2                 20           0
## 5       03/04/2020          5                 22           0
## 6       03/05/2020          3                 20           0
##   CASE_COUNT_7DAY_AVG INCOMPLETE
```

```
## 1                   NA         NA
## 2                   NA         NA
## 3                   NA         NA
## 4                   NA         NA
## 5                   NA         NA
## 6                   NA         NA
##
## Statistical summary of the data:
##  DATE_OF_INTEREST     CASE_COUNT   HOSPITALIZED_COUNT  DEATH_COUNT
##  Length:173         Min.   :   0   Min.   :   0.0     Min.   :  0.0
##  Class :character   1st Qu.: 260   1st Qu.:  31.0     1st Qu.:  8.0
##  Mode  :character   Median : 437   Median :  65.0     Median : 25.0
##                     Mean   :1316   Mean   : 320.2     Mean   :109.6
##                     3rd Qu.:2024   3rd Qu.: 359.0     3rd Qu.:131.0
##                     Max.   :6377   Max.   :1723.0     Max.   :598.0
##
##  CASE_COUNT_7DAY_AVG   INCOMPLETE
##  Min.   :   3        Min.   :6000
##  1st Qu.: 314        1st Qu.:6000
##  Median : 446        Median :6000
##  Mean   :1362        Mean   :6000
##  3rd Qu.:2030        3rd Qu.:6000
##  Max.   :5306        Max.   :6000
##  NA's   :6           NA's   :166
##
## Structure of the data:
## 'data.frame':    173 obs. of  6 variables:
##  $ DATE_OF_INTEREST   : chr  "02/29/2020" "03/01/2020" "03/02/2020" "03/03/2020" ...
##  $ CASE_COUNT         : int  1 0 0 2 5 3 8 7 21 58 ...
##  $ HOSPITALIZED_COUNT : int  11 4 21 20 22 20 18 15 17 49 ...
##  $ DEATH_COUNT        : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ CASE_COUNT_7DAY_AVG: int  NA NA NA NA NA NA 3 4 7 15 ...
##  $ INCOMPLETE         : int  NA NA NA NA NA NA NA NA NA NA ...
##
## Dimensions of the data:
## [1] 173   6
##
## Column names in the data:
## [1] "DATE_OF_INTEREST"    "CASE_COUNT"          "HOSPITALIZED_COUNT"
## [4] "DEATH_COUNT"         "CASE_COUNT_7DAY_AVG" "INCOMPLETE"
```

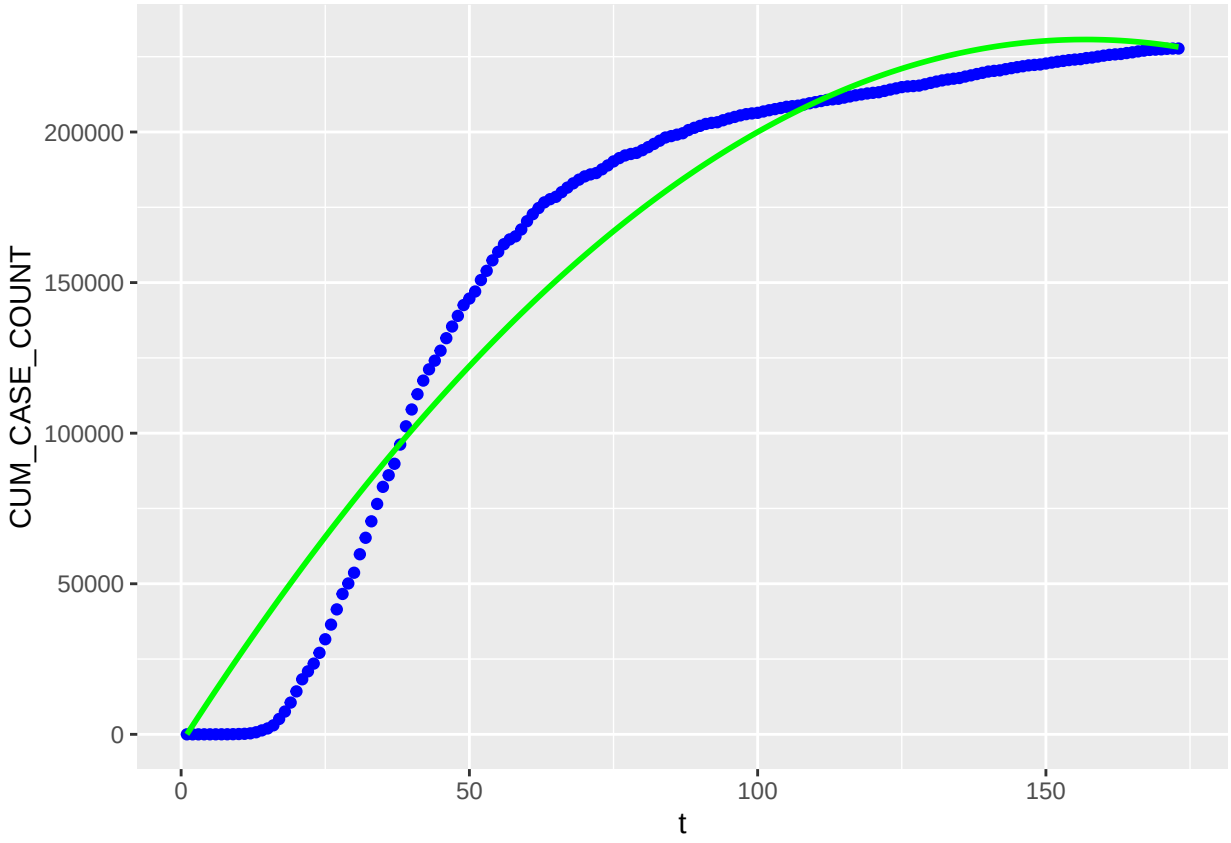
```
df$t <- c(1:dim(df)[1])
df$tsquared <- df$t^2
df$CUM_CASE_COUNT <- cumsum(df$CASE_COUNT)
```

```
m1a <- glm(CUM_CASE_COUNT ~ t + tsquared, family=poisson(link="identity"), start=c(1,1,1), data=df)
summary(m1a)
```

```
##
## Call:
## glm(formula = CUM_CASE_COUNT ~ t + tsquared, family = poisson(link = "identity"),
##     data = df, start = c(1, 1, 1))
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept) -2.969e+03  1.179e+00   -2518   <2e-16 ***
## t            2.979e+03  1.180e+00    2525   <2e-16 ***
## tsquared    -9.496e+00  9.035e-03   -1051   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 9737173  on 172  degrees of freedom
## Residual deviance: 1222503  on 170  degrees of freedom
## AIC: 1224763
##
## Number of Fisher Scoring iterations: 25
```

```
ypredm1a <- predict(m1a)

ggplot(df, aes(x = t, y = CUM_CASE_COUNT)) +
  geom_point(color='blue') +
  geom_line(aes(x = t, y = ypredm1a), size = 1, color='green')
```

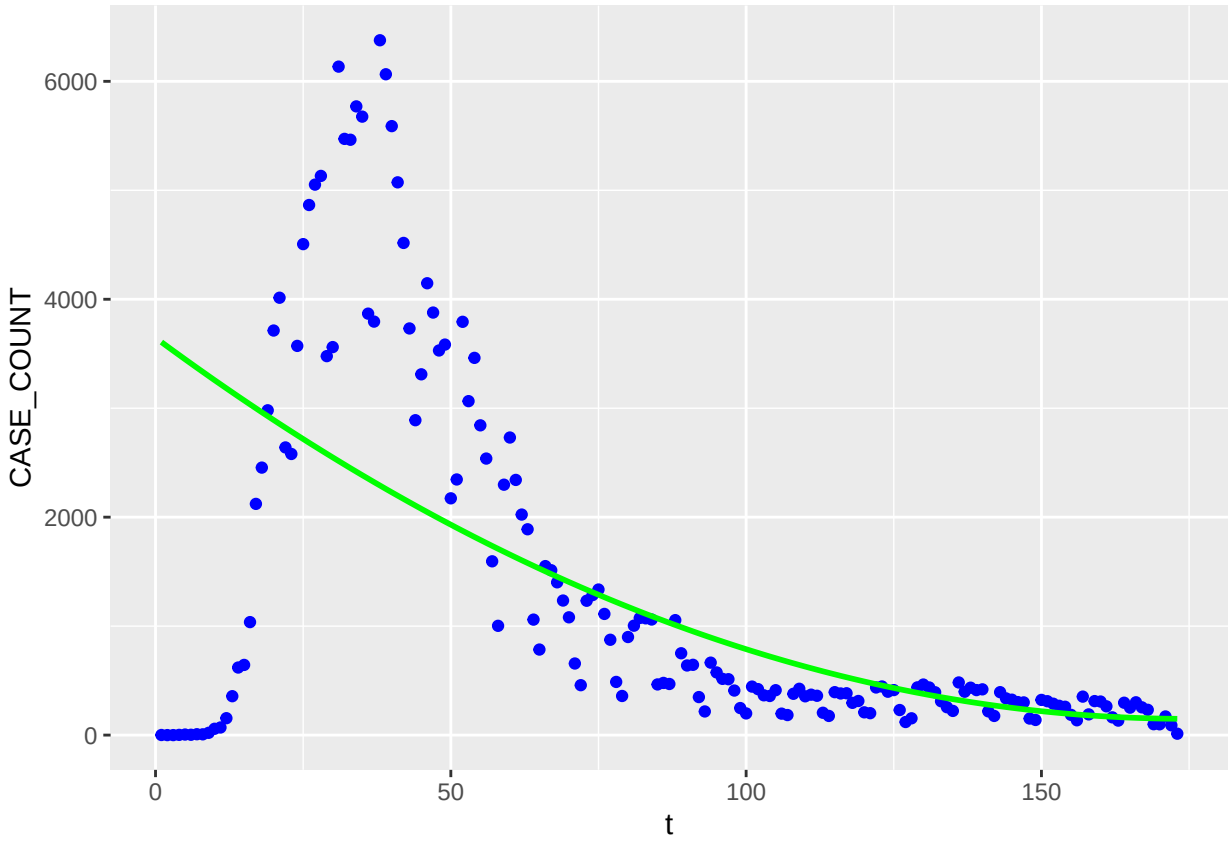



```
m1b <- glm(CASE_COUNT ~ t + tsquared, family=poisson(link="identity"), start=c(1,1,1), data=df)
summary(m1b)
```

```
##
## Call:
## glm(formula = CASE_COUNT ~ t + tsquared, family = poisson(link = "identity"),
##     data = df, start = c(1, 1, 1))
##
```

```
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.647e+03  1.086e+01   335.6   <2e-16 ***
## t           -4.004e+01  2.135e-01  -187.5   <2e-16 ***
## tsquared     1.146e-01  9.848e-04   116.4   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 292730  on 172  degrees of freedom
## Residual deviance: 159949  on 170  degrees of freedom
## AIC: 161352
##
## Number of Fisher Scoring iterations: 21
```

```
ypredm1b <- predict(m1b)

ggplot(df, aes(x = t, y = CASE_COUNT)) +
  geom_point(color='blue') +
  geom_line(aes(x = t, y = ypredm1b), size = 1, color='green')
```

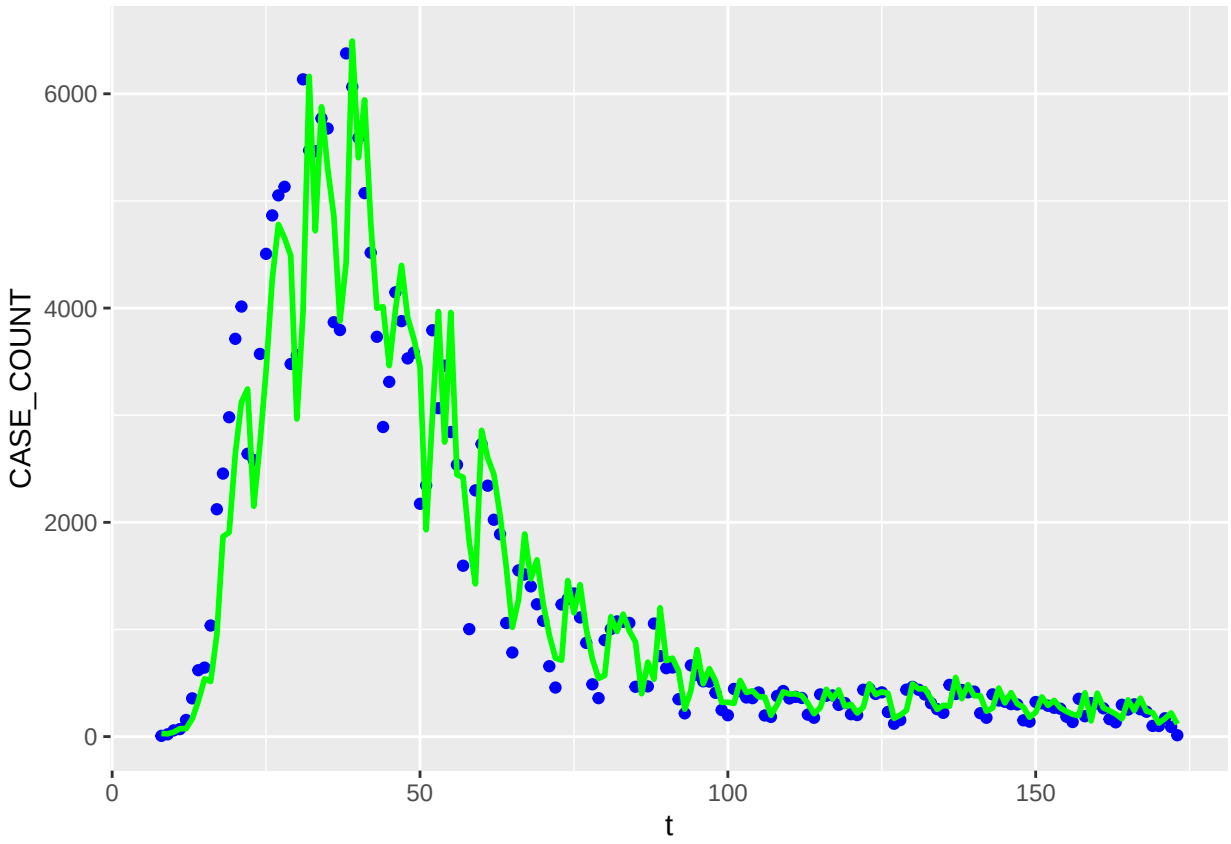



```
shift <- function(x, n){
  c(rep(NA, n), x[seq(length(x) - n)])
}

for (k in c(1:7))
{
```

```
  colname <- paste0('CASE_MINUS_', k)
  df[colname]<- shift(df$CASE_COUNT, k)
}

m2a <- glm(CASE_COUNT ~ CASE_MINUS_1 + CASE_MINUS_2 +
             CASE_MINUS_3 + CASE_MINUS_4 + CASE_MINUS_5 +
             CASE_MINUS_6 + CASE_MINUS_7, family=poisson(link="identity"),
              start=c(1,1,1,1,1,1,1,1), data=df[8:nrow(df),])
summary(m2a)
```

```
##
## Call:
## glm(formula = CASE_COUNT ~ CASE_MINUS_1 + CASE_MINUS_2 + CASE_MINUS_3 +
##     CASE_MINUS_4 + CASE_MINUS_5 + CASE_MINUS_6 + CASE_MINUS_7,
##     family = poisson(link = "identity"), data = df[8:nrow(df),
##         ], start = c(1, 1, 1, 1, 1, 1, 1, 1))
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  21.445623   1.704655  12.581  < 2e-16 ***
## CASE_MINUS_1  1.007670   0.008443 119.344  < 2e-16 ***
## CASE_MINUS_2 -0.401347   0.011247 -35.684  < 2e-16 ***
## CASE_MINUS_3  0.282205   0.011742  24.034  < 2e-16 ***
## CASE_MINUS_4 -0.108602   0.011761  -9.234  < 2e-16 ***
## CASE_MINUS_5 -0.092772   0.010716  -8.658  < 2e-16 ***
## CASE_MINUS_6  0.251908   0.009984  25.231  < 2e-16 ***
## CASE_MINUS_7  0.046078   0.007575   6.083 1.18e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 274172  on 165  degrees of freedom
## Residual deviance:  15619  on 158  degrees of freedom
## AIC: 17017
##
## Number of Fisher Scoring iterations: 9
```

```
ypredm2a <- predict(m2a)

ggplot(df[8:nrow(df),], aes(x = t, y = CASE_COUNT)) +
  geom_point(color='blue') +
  geom_line(aes(x = t, y = ypredm2a), size = 1, color='green')
```

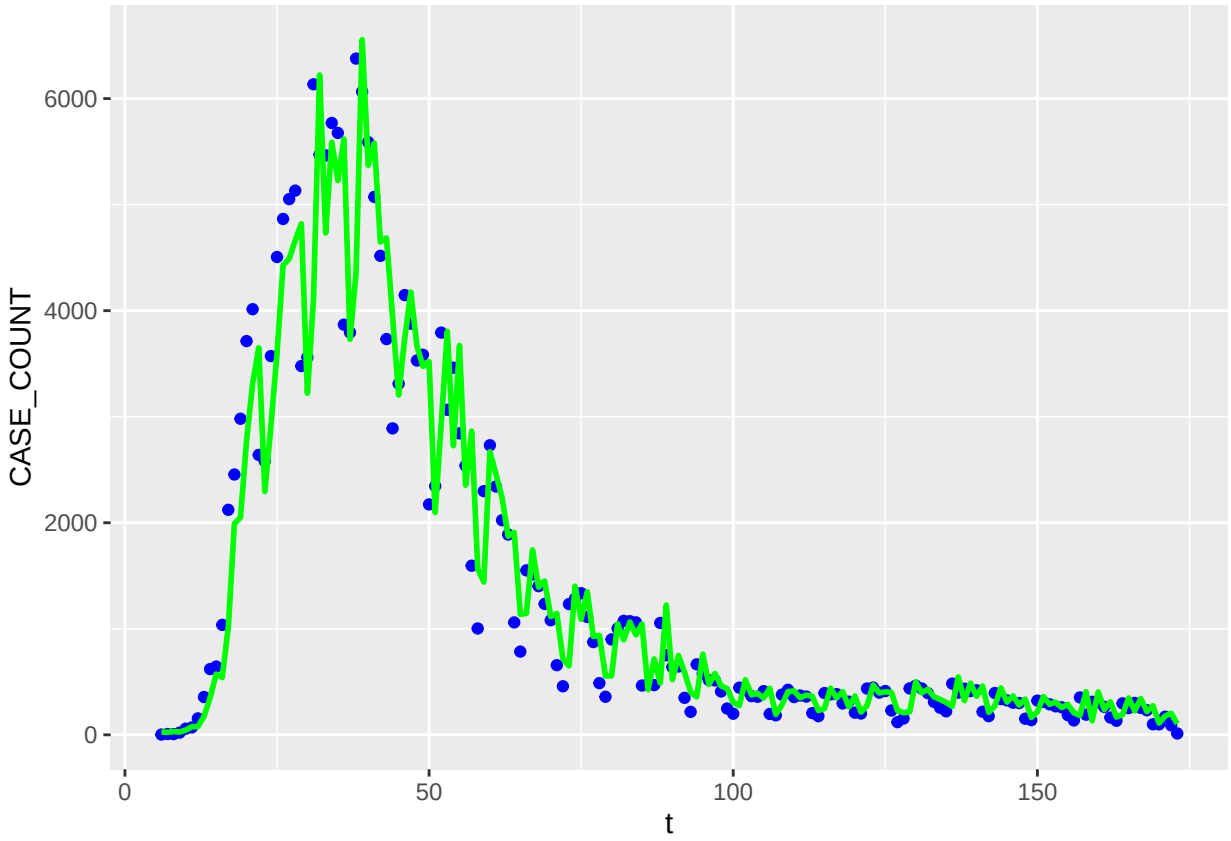



```
m2a5 <- glm(CASE_COUNT ~ CASE_MINUS_1 + CASE_MINUS_2 +
              CASE_MINUS_3 + CASE_MINUS_4 + CASE_MINUS_5, family=poisson(link="identity"),
            start=c(1,1,1,1,1,1), data=df[6:nrow(df),])

summary(m2a5)
```

```
##
## Call:
## glm(formula = CASE_COUNT ~ CASE_MINUS_1 + CASE_MINUS_2 + CASE_MINUS_3 +
##     CASE_MINUS_4 + CASE_MINUS_5, family = poisson(link = "identity"),
##     data = df[6:nrow(df), ], start = c(1, 1, 1, 1, 1, 1))
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  22.302889   1.543974   14.45   <2e-16 ***
## CASE_MINUS_1  1.056281   0.007836  134.80   <2e-16 ***
## CASE_MINUS_2 -0.441629   0.010965  -40.27   <2e-16 ***
## CASE_MINUS_3  0.354777   0.011369   31.20   <2e-16 ***
## CASE_MINUS_4 -0.182627   0.011309  -16.15   <2e-16 ***
## CASE_MINUS_5  0.197119   0.007736   25.48   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 279485  on 167  degrees of freedom
## Residual deviance:  16954  on 162  degrees of freedom
```

```
## AIC: 18355
##
## Number of Fisher Scoring iterations: 7
```

```
ypredm2a5 <- predict(m2a5)

ggplot(df[6:nrow(df),], aes(x = t, y = CASE_COUNT)) +
  geom_point(color='blue') +
  geom_line(aes(x = t, y = ypredm2a5), size = 1, color='green')
```

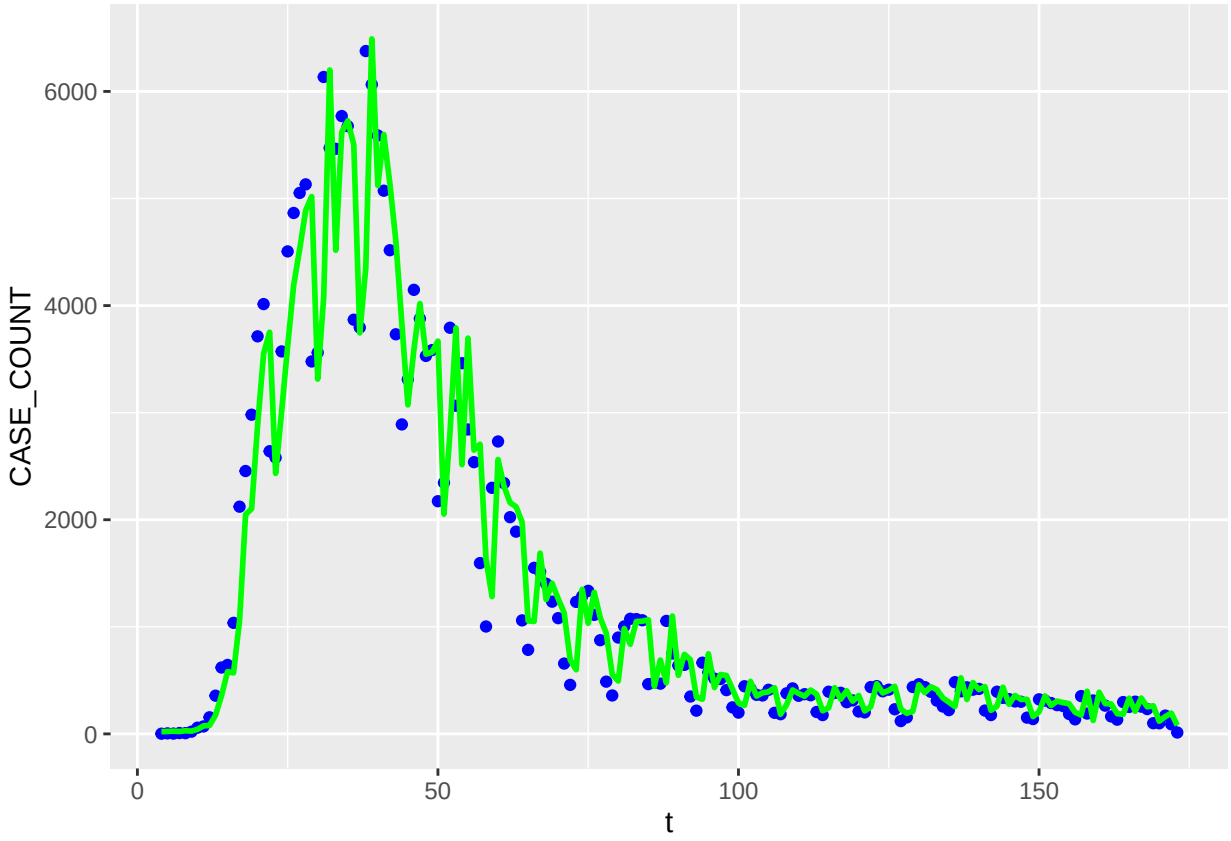



```
m2a3 <- glm(CASE_COUNT ~ CASE_MINUS_1 + CASE_MINUS_2 +
              CASE_MINUS_3, family=poisson(link="identity"),
            start=c(1,1,1,1), data=df[4:nrow(df),])

summary(m2a3)
```

```
##
## Call:
## glm(formula = CASE_COUNT ~ CASE_MINUS_1 + CASE_MINUS_2 + CASE_MINUS_3,
##     family = poisson(link = "identity"), data = df[4:nrow(df),
##         ], start = c(1, 1, 1, 1))
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  20.673739   1.375165   15.03   <2e-16 ***
## CASE_MINUS_1  1.049060   0.007662  136.92   <2e-16 ***
## CASE_MINUS_2 -0.394642   0.010492  -37.61   <2e-16 ***
## CASE_MINUS_3  0.330427   0.007205   45.86   <2e-16 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 284779  on 169  degrees of freedom
## Residual deviance:  17668  on 166  degrees of freedom
## AIC: 19071
##
## Number of Fisher Scoring iterations: 8
```

```
ypredm2a3 <- predict(m2a3)

ggplot(df[4:nrow(df),], aes(x = t, y = CASE_COUNT)) +
  geom_point(color='blue') +
  geom_line(aes(x = t, y = ypredm2a3), size = 1, color='green')
```




```
#install.packages("readxl")
```

```
library(readxl)
esti_asymptomatic <- read.csv('esti_asymptomatic.csv')
names(esti_asymptomatic)[names(esti_asymptomatic) == "P.Sev.1...age."] <- "P_Sev_age"
esti_asymptomatic
```

```
##      Age     Male   Female           X. P_Sev_age Population_perc   PDF.asymp.
## 1    0-4 45106267 39934767 5.931232e-02       27%    1.590998e-02 3.162801e-02
## 2   5-10 46319087 40225712 6.036113e-02       27%    1.619132e-02 3.218728e-02
## 3  11-14 45044032 38648113 5.837153e-02       27%    1.565762e-02 3.112634e-02
```

```
## 4  15-19 44286173 38535391 5.776434e-02       51%    2.953473e-02 5.871312e-02
## 5  20-24 46956410 41620012 6.177809e-02       75%    4.660250e-02 9.264274e-02
## 6  25-29 54633535 49681080 7.275478e-02       75%    5.488280e-02 1.091034e-01
## 7  30-34 65544747 61670474 8.872693e-02       75%    6.693144e-02 1.330553e-01
## 8  35-39 49370823 46920301 6.715875e-02       75%    5.066142e-02 1.007116e-01
## 9  40-44 51155493 48766184 6.969090e-02       75%    5.257155e-02 1.045088e-01
## 10 45-49 62836913 60442592 8.598194e-02       48%    4.135601e-02 8.221305e-02
## 11 50-54 60946291 59710325 8.415259e-02       48%    4.047612e-02 8.046389e-02
## 12 55-59 47588180 46379016 6.553792e-02       48%    3.152274e-02 6.266516e-02
## 13 60-64 38607054 38368062 5.368670e-02       48%    2.582249e-02 5.133342e-02
## 14 65-69 35185041 36260434 4.983002e-02       13%    6.477903e-03 1.287765e-02
## 15 70-74 19746314 21631501 2.885918e-02       13%    3.751693e-03 7.458121e-03
## 16 75-79 11801212 13715744 1.779694e-02       13%    2.313602e-03 4.599289e-03
## 17 80-84  6808650  9097473 1.109381e-02       13%    1.442195e-03 2.866990e-03
## 18 85-89  2787103  4583542 5.140695e-03       13%    6.682904e-04 1.328518e-03
## 19 90-94   748881  1525990 1.586621e-03       13%    2.062607e-04 4.100328e-04
## 20 95-99   140322   386612 3.675129e-04       13%    4.777668e-05 9.497692e-05
## 21  100+    11731    56108 4.731467e-05       13%    6.150907e-06 1.222760e-05
##    CDF.asymp.
## 1  0.03162801
## 2  0.06381529
## 3  0.09494162
## 4  0.15365474
## 5  0.24629748
## 6  0.35540091
## 7  0.48845625
## 8  0.58916785
## 9  0.69367667
## 10 0.77588972
## 11 0.85635361
## 12 0.91901877
## 13 0.97035220
## 14 0.98322985
## 15 0.99068797
## 16 0.99528725
## 17 0.99815425
## 18 0.99948276
## 19 0.99989280
## 20 0.99998777
## 21 1.00000000
```

```r
categorize_age <- function(age) {
  if (age %in% c("0-4", "5-10", "11-14")) return("0-17")
  if (age %in% c("15-19", "20-24", "25-29", "30-34", "35-39", "40-44")) return("18-44")
  if (age %in% c("45-49", "50-54", "55-59", "60-64")) return("45-64")
  if (age %in% c("65-69", "70-74")) return("65-74")
  return("75+")
}

esti_asymptomatic$Broad_Age_Group <- sapply(esti_asymptomatic$Age, categorize_age)

assign_severity <- function(broad_age) {
  switch(broad_age,
         "0-17" = 0.27,
```

```
         "18-44" = 0.75,
         "45-64" = 0.48,
         "65-74" = 0.13,
         "75+" = 0.13)
}

esti_asymptomatic$Assigned_Sev_Percentage <- sapply(esti_asymptomatic$Broad_Age_Group, assign_severity)

summary_data <- unique(esti_asymptomatic[, c("Broad_Age_Group", "Assigned_Sev_Percentage")])
colnames(summary_data) <- c("Age Group", "Sev 1 | age")

print(summary_data)
```

```
##    Age Group Sev 1 | age
## 1       0-17        0.27
## 4      18-44        0.75
## 10     45-64        0.48
## 14     65-74        0.13
## 16       75+        0.13
```

```
#Calculate the total population for each broad age group:
esti_asymptomatic$Total_Population <- esti_asymptomatic$Male + esti_asymptomatic$Female
age_group_population <- aggregate(Total_Population ~ Broad_Age_Group, data = esti_asymptomatic, FUN = su
total_population <- sum(esti_asymptomatic$Total_Population)
#Calculate the total asymptomatic individuals for each broad age group:
esti_asymptomatic$Asymptomatic_Count <- esti_asymptomatic$PDF.asymp. * esti_asymptomatic$Total_Populati
age_group_asymptomatic <- aggregate(Asymptomatic_Count ~ Broad_Age_Group, data = esti_asymptomatic, FUN
#Identify which broad age group has the highest asymptomatic proportion:
age_group_asymptomatic$Asymptomatic_Proportion <- age_group_asymptomatic$Asymptomatic_Count / total_popu
highest_asymptomatic_group <- age_group_asymptomatic[which.max(age_group_asymptomatic$Asymptomatic_Prop
```

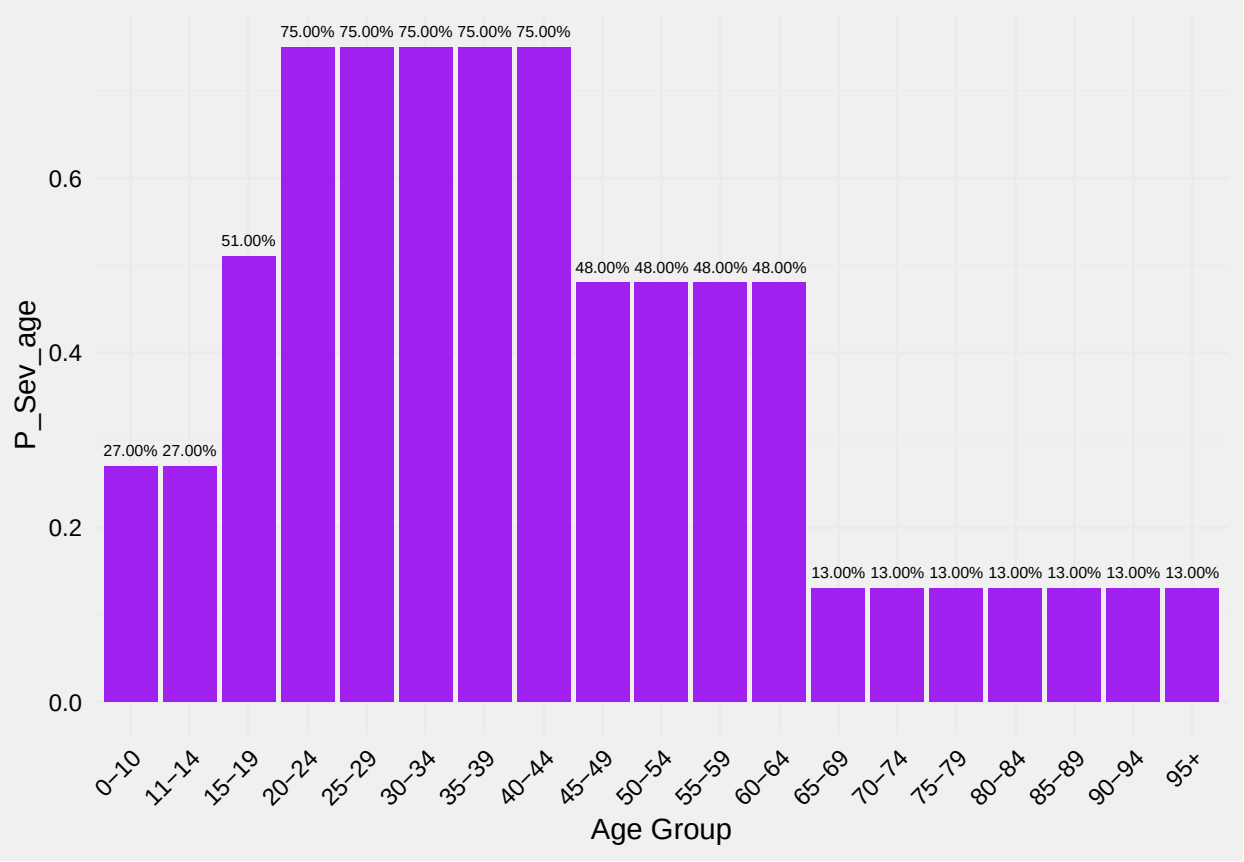
```
merged_data <- merge(age_group_population, age_group_asymptomatic, by = "Broad_Age_Group")

overall_data <- data.frame(
  Broad_Age_Group = "Overall",
  Total_Population = total_population,
  Asymptomatic_Count = sum(merged_data$Asymptomatic_Count),
  Asymptomatic_Proportion = sum(merged_data$Asymptomatic_Count) / total_population
)
highest_group_data <- data.frame(
  Broad_Age_Group = "Highest Asymptomatic Group",
  Total_Population = NA,
  Asymptomatic_Count = NA,
  Asymptomatic_Proportion = max(merged_data$Asymptomatic_Proportion)
)

final_data <- rbind(merged_data, overall_data, highest_group_data)

print(final_data)
```

```
##              Broad_Age_Group Total_Population Asymptomatic_Count
## 1                       0-17        255277978          8080350.0
## 2                      18-44        599140623         61516750.8
## 3                      45-64        414878433         29683550.9
## 4                      65-74        112823290          1228650.5
```

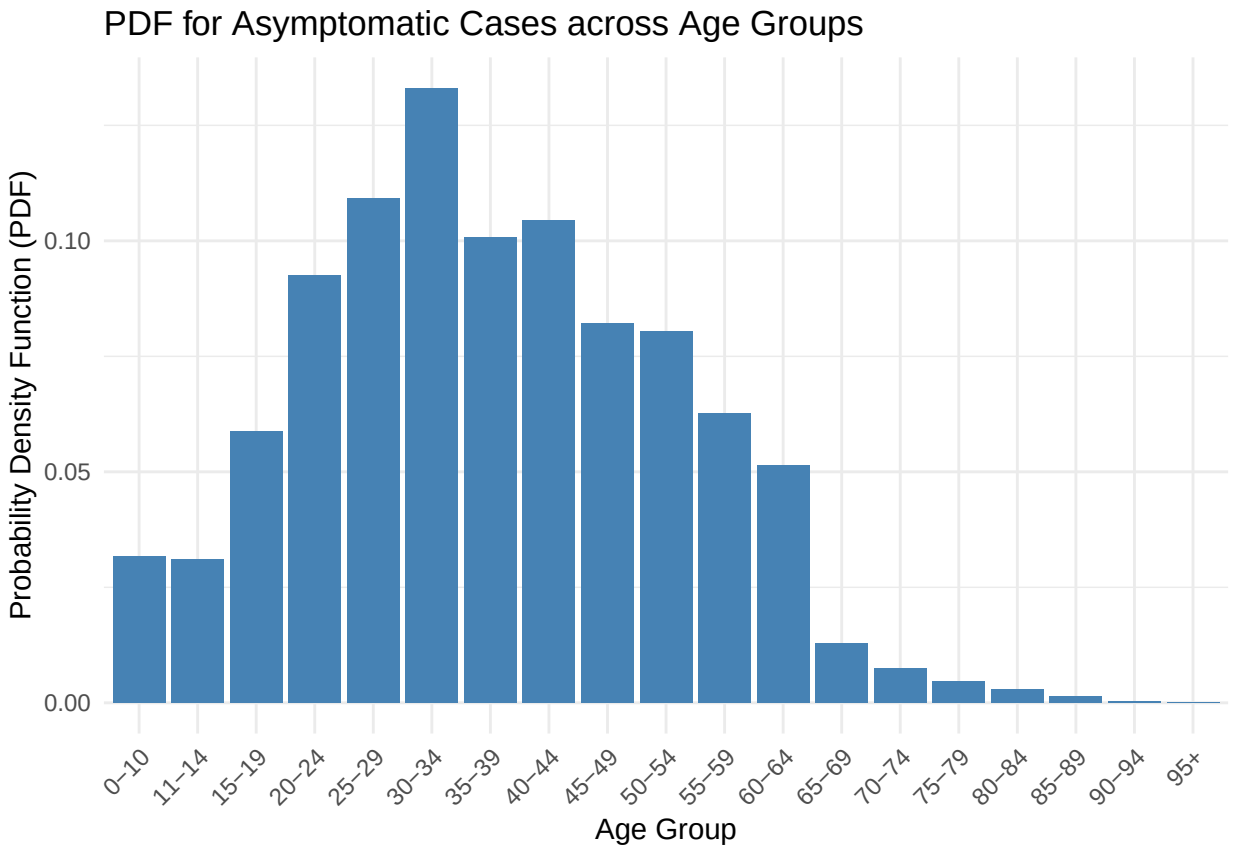
```
## 5                         75+         51663368           173738.2
## 6                     Overall       1433783692        100683040.4
## 7 Highest Asymptomatic Group               NA                 NA
##   Asymptomatic_Proportion
## 1              0.0056356828
## 2              0.0429051824
## 3              0.0207029492
## 4              0.0008569288
## 5              0.0001211746
## 6              0.0702219177
## 7              0.0429051824
```

```r
library(ggplot2)
library(dplyr)
library(scales)
```

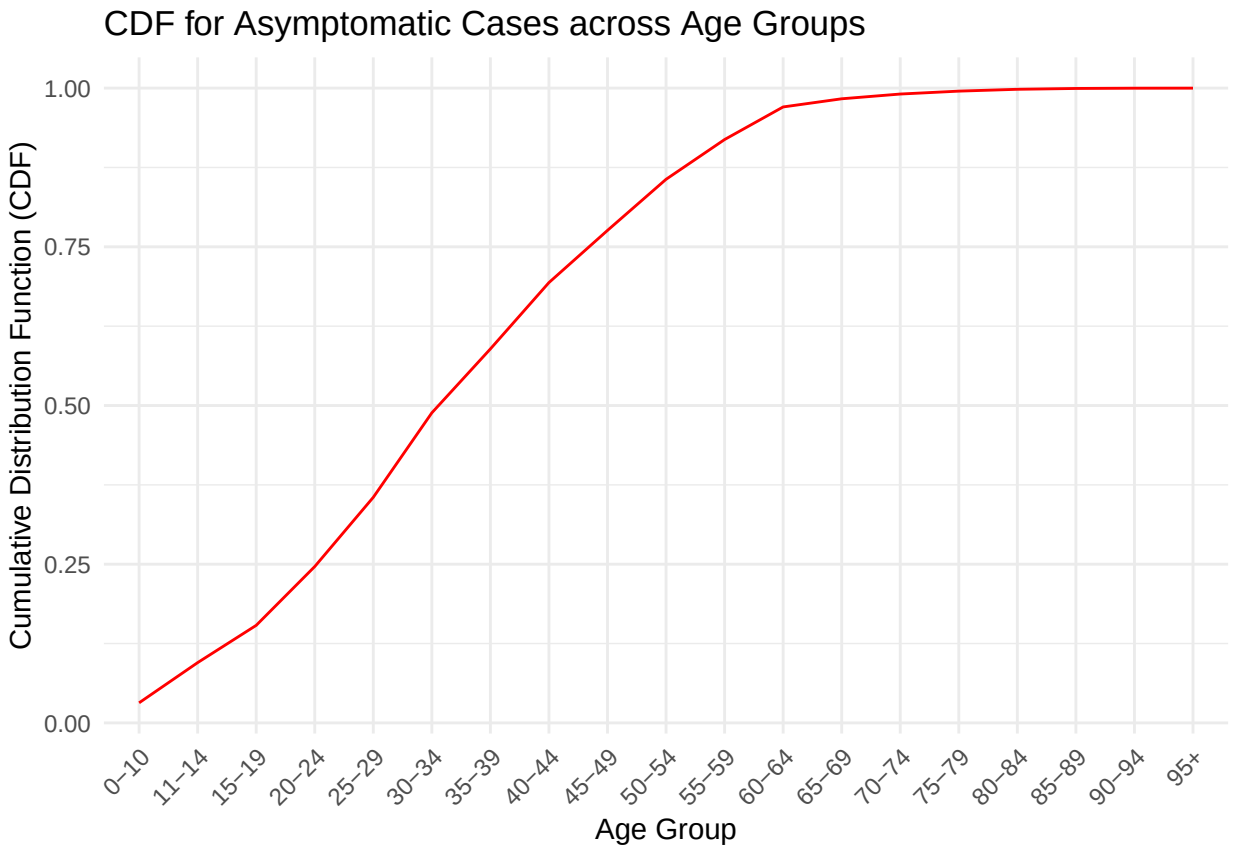
```r
esti_asymptomatic$P_Sev_age <- as.numeric(gsub("%", "", esti_asymptomatic$P_Sev_age)) / 100
esti_asymptomatic <- esti_asymptomatic[!(esti_asymptomatic$Age %in% c("5-10", "100+")), ]
esti_asymptomatic$Age[esti_asymptomatic$Age == "0-4"] <- "0-10"
esti_asymptomatic$Age[esti_asymptomatic$Age == "95-99"] <- "95+"


# Plotting the distribution of P_Sev_age
ggplot(data = esti_asymptomatic, aes(x = Age, y = P_Sev_age)) +
  geom_bar(stat = "identity", fill = "purple") +
  theme_minimal() +
  theme(
    plot.background = element_rect(fill = "#F0F0F0", color = NA),  # Light purple background for the en
    axis.text.x = element_text(angle = 45, hjust = 1),
    axis.text = element_text(color = "black")  # Ensuring text is visible on the light background
  ) +
  labs(
    #title = "Distribution of P_Sev_age across Age Groups",
    x = "Age Group",
    y = "P_Sev_age"
  ) +
  geom_text(
    aes(label = scales::percent(P_Sev_age, accuracy = 0.01)),
    vjust = -0.9,
    size = 2
  )
```



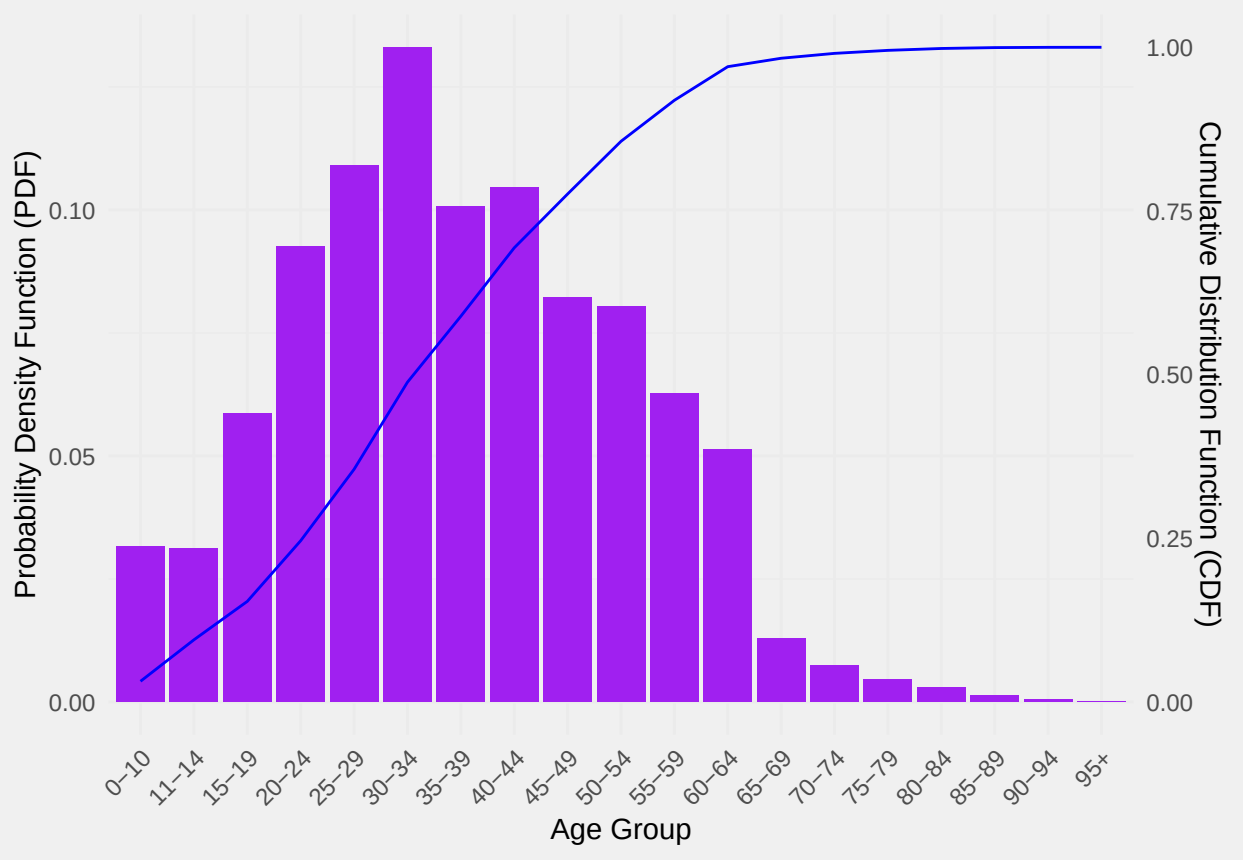

```r
# Plotting PDF for asymptomatic cases
ggplot(data = esti_asymptomatic, aes(x = Age, y = PDF.asymp.)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  theme_minimal() +
  labs(title = "PDF for Asymptomatic Cases across Age Groups",
       x = "Age Group",
       y = "Probability Density Function (PDF)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
# Plotting CDF for asymptomatic cases
ggplot(data = esti_asymptomatic, aes(x = Age, y = CDF.asymp., group = 1)) +
  geom_line(color = "red") +
  theme_minimal() +
  labs(title = "CDF for Asymptomatic Cases across Age Groups",
       x = "Age Group",
       y = "Cumulative Distribution Function (CDF)") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
scaling_factor <- max(esti_asymptomatic$PDF.asymp.) / max(esti_asymptomatic$CDF.asymp.)

ggplot(data = esti_asymptomatic, aes(x = Age)) +
  # Plot PDF.asymp.
  geom_bar(aes(y = PDF.asymp.), stat = "identity", fill = "purple") +

  # Plot CDF.asymp. scaled by the factor
  geom_line(aes(y = CDF.asymp. * scaling_factor, group = 1), color = "blue") +

  labs(#title = "PDF & CDF for Asymptomatic Cases across Age Groups",
       x = "Age Group",
       y = "Probability Density Function (PDF)") +

  scale_y_continuous(sec.axis = sec_axis(~./scaling_factor, name = "Cumulative Distribution Function (CD

  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.background = element_rect(fill = "#F0F0F0", color = NA),)
```

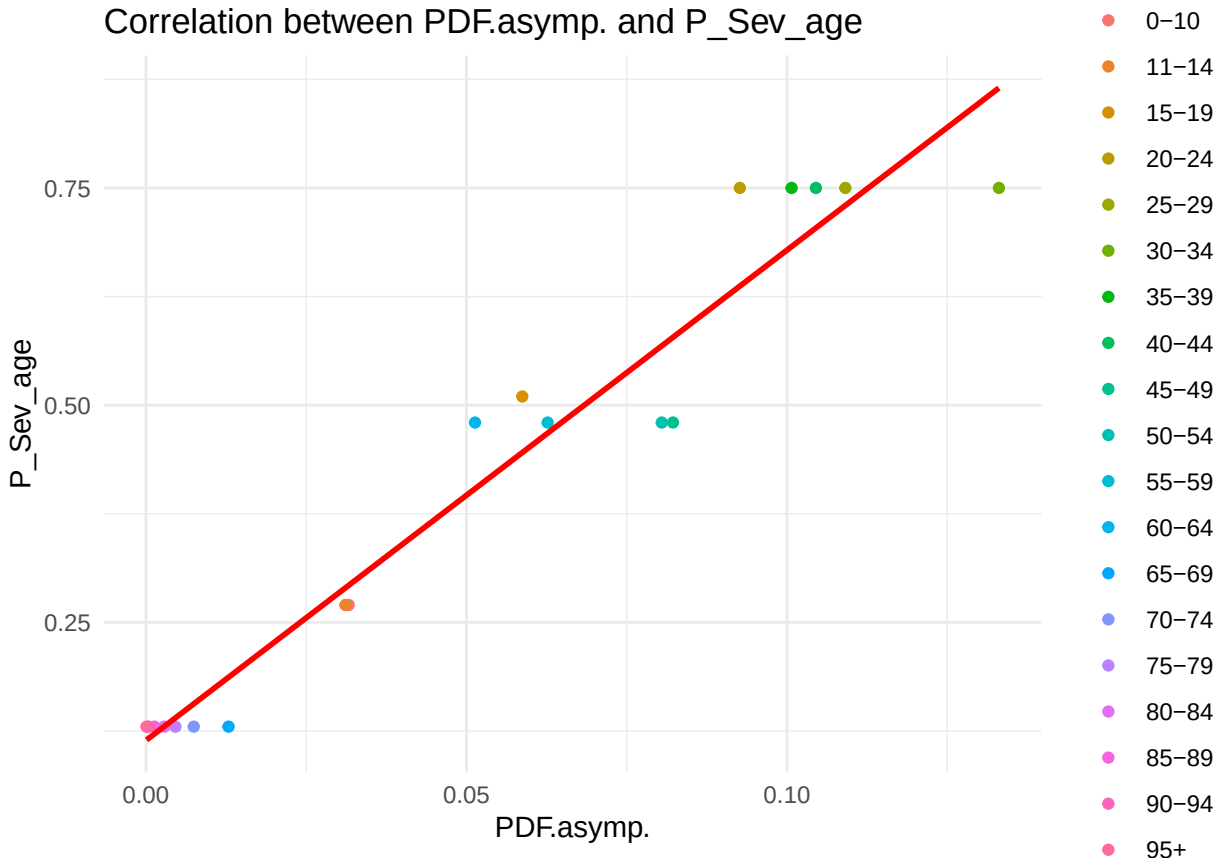



```
#Calculate the total probability of being asymptomatic in the entire population.
esti_asymptomatic$Weighted_PDF <- esti_asymptomatic$PDF.asymp. * esti_asymptomatic$Total_Population
total_probability_asymptomatic <- sum(esti_asymptomatic$Weighted_PDF) / total_population
```

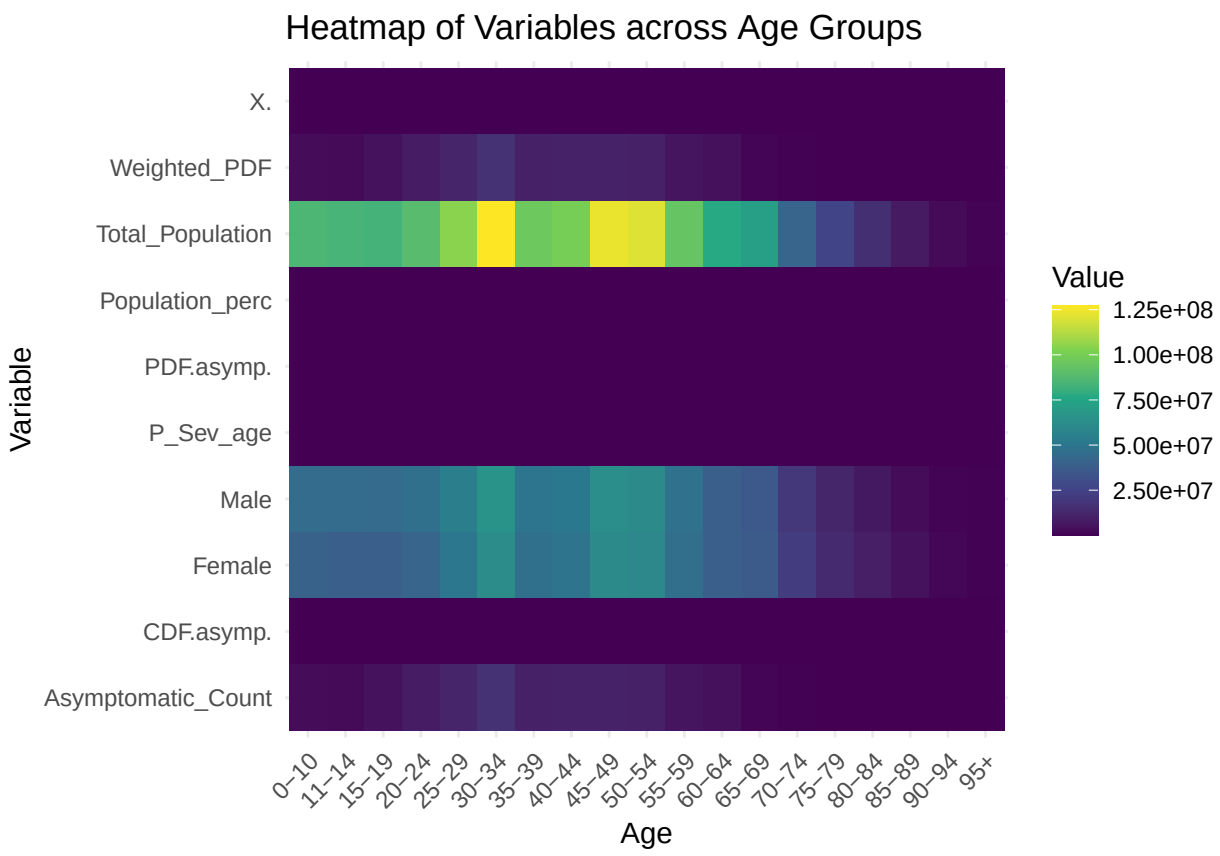
```
#Determine the correlation between different variables, e.g., how PDF.asymp. correlates with P_Sev_age
correlation <- cor(esti_asymptomatic$PDF.asymp., esti_asymptomatic$P_Sev_age)
print(paste("The correlation between PDF.asymp. and P_Sev_age is:", round(correlation, 2)))
```

```
## [1] "The correlation between PDF.asymp. and P_Sev_age is: 0.97"
```

```
ggplot(data = esti_asymptomatic, aes(x = PDF.asymp., y = P_Sev_age)) +
  geom_point(aes(color = Age)) +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  theme_minimal() +
  labs(title = "Correlation between PDF.asymp. and P_Sev_age",
       x = "PDF.asymp.",
       y = "P_Sev_age")
```




```r
library(ggplot2)
library(tidyr)

long_data <- gather(esti_asymptomatic, key = "Variable", value = "Value", -Age, -Broad_Age_Group, -Assig

ggplot(long_data, aes(x = Age, y = Variable, fill = Value)) +
  geom_tile() +
  scale_fill_viridis_c() +
  labs(title = "Heatmap of Variables across Age Groups") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```


Heatmap of Variables across Age Groups
Variable
X.
Weighted_PDF
Total_Population
Population_perc
PDF.asymp.
P_Sev_age
Male
Female
CDF.asymp.
Asymptomatic_Count
0–10
11–14
15–19
20–24
25–29
30–34
35–39
40–44
45–49
50–54
55–59
60–64
65–69
70–74
75–79
80–84
85–89
90–94
95+
Age
Value
1.25e+08
1.00e+08
7.50e+07
5.00e+07
2.50e+07