

# A Descriptive Analysis of County-level Confirmed COVID-19 Cases in U.S.

NYU A3SR

```
if(!requireNamespace("dplyr"))  
  install.packages("dplyr", repos = "https://cloud.r-project.org")  
if(!requireNamespace("data.table"))  
  install.packages("data.table", repos = "https://cloud.r-project.org")  
if(!requireNamespace("here"))  
  install.packages("here", repos = "https://cloud.r-project.org")  
if(!requireNamespace("ggplot2"))  
  install.packages("ggplot2", repos = "https://cloud.r-project.org")  
require(data.table)  
require(dplyr)  
library(here)  
require(ggplot2)
```

## Introduction

This descriptive analysis illustrates how county-level confirmed COVID-19 cases were developed in the United States, starting from January, 2020. The data we used in this study come from USA FACTS. The data files are updated on a daily basis and are stored in time stamp formats. We applied multiple statistical tools to analyze and visualize the outbreak development.

## Creating Data Files

Through data processing, we created two data files for the analysis:

1. Daily nation-wide total confirmed cases
2. Most recent confirmed cases by county (as of 03-28-2020)

## Descriptive Analysis

### Data Structure

Data are ordered in time series. Starting from Jan. 22, 2020 (1/22/2020), county-level confirmed cases were recorded daily. Each value time stamp variables is cumulative. For example, the value of confirmed cases in Queens, NY on Mar. 28, 2020 is calculated by adding cumulative confirmed cases on Mar. 27, 2020 with new cases confirmed on Mar. 28. Therefore, to get daily new cases, we can simply subtract the value on a particular day with the value in the previous day.

## Overview Top 5 rows of raw data

```
head(dat, 5)
```

```
##      countyFIPS      County Name State stateFIPS 1/22/2020 1/23/2020
## 1:         0 Statewide Unallocated   AL         1         0         0
## 2:        1001      Autauga County   AL         1         0         0
## 3:        1003      Baldwin County   AL         1         0         0
## 4:        1005      Barbour County   AL         1         0         0
## 5:        1007        Bibb County   AL         1         0         0
## 1/24/2020 1/25/2020 1/26/2020 1/27/2020 1/28/2020 1/29/2020 1/30/2020
## 1:         0         0         0         0         0         0         0
## 2:         0         0         0         0         0         0         0
## 3:         0         0         0         0         0         0         0
## 4:         0         0         0         0         0         0         0
## 5:         0         0         0         0         0         0         0
## 1/31/2020 2/1/2020 2/2/2020 2/3/2020 2/4/2020 2/5/2020 2/6/2020 2/7/2020
## 1:         0         0         0         0         0         0         0         0
## 2:         0         0         0         0         0         0         0         0
## 3:         0         0         0         0         0         0         0         0
## 4:         0         0         0         0         0         0         0         0
## 5:         0         0         0         0         0         0         0         0
## 2/8/2020 2/9/2020 2/10/2020 2/11/2020 2/12/2020 2/13/2020 2/14/2020
## 1:         0         0         0         0         0         0         0         0
## 2:         0         0         0         0         0         0         0         0
## 3:         0         0         0         0         0         0         0         0
## 4:         0         0         0         0         0         0         0         0
## 5:         0         0         0         0         0         0         0         0
## 2/15/2020 2/16/2020 2/17/2020 2/18/2020 2/19/2020 2/20/2020 2/21/2020
## 1:         0         0         0         0         0         0         0         0
## 2:         0         0         0         0         0         0         0         0
## 3:         0         0         0         0         0         0         0         0
## 4:         0         0         0         0         0         0         0         0
## 5:         0         0         0         0         0         0         0         0
## 2/22/2020 2/23/2020 2/24/2020 2/25/2020 2/26/2020 2/27/2020 2/28/2020
## 1:         0         0         0         0         0         0         0         0
## 2:         0         0         0         0         0         0         0         0
## 3:         0         0         0         0         0         0         0         0
## 4:         0         0         0         0         0         0         0         0
## 5:         0         0         0         0         0         0         0         0
## 2/29/2020 3/1/2020 3/2/2020 3/3/2020 3/4/2020 3/5/2020 3/6/2020 3/7/2020
## 1:         0         0         0         0         0         0         0         0
## 2:         0         0         0         0         0         0         0         0
## 3:         0         0         0         0         0         0         0         0
## 4:         0         0         0         0         0         0         0         0
## 5:         0         0         0         0         0         0         0         0
## 3/8/2020 3/9/2020 3/10/2020 3/11/2020 3/12/2020 3/13/2020 3/14/2020
## 1:         0         0         0         0         0         0         0         0
## 2:         0         0         0         0         0         0         0         0
## 3:         0         0         0         0         0         0         1         0
## 4:         0         0         0         0         0         0         0         0
## 5:         0         0         0         0         0         0         0         0
## 3/15/2020 3/16/2020 3/17/2020 3/18/2020 3/19/2020 3/20/2020 3/21/2020
## 1:         0         0         0         0         0         0         0         0
```

```
## 2:      0      0      0      0      0      0      0
## 3:      1      1      1      1      1      2      2
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
##    3/22/2020 3/23/2020 3/24/2020 3/25/2020 3/26/2020 3/27/2020 3/28/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      1      4      6      6      6
## 3:      3      3      4      4      5      5     10
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
```

Bottom 5 rows of raw data

```
tail(dat, 5)
```

```
##    countyFIPS      County Name State stateFIPS 1/22/2020 1/23/2020 1/24/2020
## 1:      56037 Sweetwater County   WY      56      0      0      0
## 2:      56039   Teton County     WY      56      0      0      0
## 3:      56041   Uinta County     WY      56      0      0      0
## 4:      56043 Washakie County     WY      56      0      0      0
## 5:      56045   Weston County     WY      56      0      0      0
##    1/25/2020 1/26/2020 1/27/2020 1/28/2020 1/29/2020 1/30/2020 1/31/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
##    2/1/2020 2/2/2020 2/3/2020 2/4/2020 2/5/2020 2/6/2020 2/7/2020 2/8/2020
## 1:      0      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0      0
##    2/9/2020 2/10/2020 2/11/2020 2/12/2020 2/13/2020 2/14/2020 2/15/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
##    2/16/2020 2/17/2020 2/18/2020 2/19/2020 2/20/2020 2/21/2020 2/22/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
##    2/23/2020 2/24/2020 2/25/2020 2/26/2020 2/27/2020 2/28/2020 2/29/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
##    3/1/2020 3/2/2020 3/3/2020 3/4/2020 3/5/2020 3/6/2020 3/7/2020 3/8/2020
## 1:      0      0      0      0      0      0      0      0
```

```
## 2:      0      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0      0
##      3/9/2020 3/10/2020 3/11/2020 3/12/2020 3/13/2020 3/14/2020 3/15/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
##      3/16/2020 3/17/2020 3/18/2020 3/19/2020 3/20/2020 3/21/2020 3/22/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      1      1      2      2      2
## 3:      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
##      3/23/2020 3/24/2020 3/25/2020 3/26/2020 3/27/2020 3/28/2020
## 1:      0      1      1      1      1      1
## 2:      2      3      6      8     12     14
## 3:      0      0      0      0      0      0
## 4:      0      0      0      0      1      1
## 5:      0      0      0      0      0      0
```

**Daily Nation-wide Total** A summary table of nation-wide daily total confirmed cases is attached.

```
# Create a table showing daily total confirmed cases, nation-wide
confirmed_cases_per_day <- dat[, 5:71]
date <- as.Date(paste("0", colnames(confirmed_cases_per_day) , sep = ""), "%m/%d/%y")

total_confirmed_daily <- lapply(dat[, 5:71], FUN = sum)
total_confirmed_daily <- t(as.matrix(as.data.frame(total_confirmed_daily)))
total_confirmed_daily <- data.frame(date = date,
                                     total_confirmed = total_confirmed_daily)
row.names(total_confirmed_daily) <- NULL

total_confirmed_daily
```

```
##      date total_confirmed
## 1 2020-01-22              1
## 2 2020-01-23              1
## 3 2020-01-24              2
## 4 2020-01-25              2
## 5 2020-01-26              5
## 6 2020-01-27              5
## 7 2020-01-28              5
## 8 2020-01-29              5
## 9 2020-01-30              5
## 10 2020-01-31             7
## 11 2020-02-01             8
## 12 2020-02-02             8
## 13 2020-02-03            11
## 14 2020-02-04            11
## 15 2020-02-05            11
```

## 16	2020-02-06	11
## 17	2020-02-07	11
## 18	2020-02-08	11
## 19	2020-02-09	11
## 20	2020-02-10	11
## 21	2020-02-11	12
## 22	2020-02-12	12
## 23	2020-02-13	13
## 24	2020-02-14	13
## 25	2020-02-15	13
## 26	2020-02-16	13
## 27	2020-02-17	13
## 28	2020-02-18	13
## 29	2020-02-19	13
## 30	2020-02-20	13
## 31	2020-02-21	15
## 32	2020-02-22	15
## 33	2020-02-23	15
## 34	2020-02-24	15
## 35	2020-02-25	15
## 36	2020-02-26	15
## 37	2020-02-27	16
## 38	2020-02-28	16
## 39	2020-02-29	24
## 40	2020-03-01	30
## 41	2020-03-02	53
## 42	2020-03-03	72
## 43	2020-03-04	103
## 44	2020-03-05	171
## 45	2020-03-06	211
## 46	2020-03-07	360
## 47	2020-03-08	477
## 48	2020-03-09	727
## 49	2020-03-10	997
## 50	2020-03-11	1325
## 51	2020-03-12	1713
## 52	2020-03-13	2242
## 53	2020-03-14	2841
## 54	2020-03-15	3643
## 55	2020-03-16	4602
## 56	2020-03-17	6113
## 57	2020-03-18	8997
## 58	2020-03-19	13975
## 59	2020-03-20	18837
## 60	2020-03-21	26184
## 61	2020-03-22	33469
## 62	2020-03-23	43544
## 63	2020-03-24	53552
## 64	2020-03-25	68245
## 65	2020-03-26	84677
## 66	2020-03-27	102269
## 67	2020-03-28	122714

```

# Create a tidy datafile to include only
# county, state, and most recent confirmed cases.
total_cases <- total_confirmed_daily[nrow(total_confirmed_daily), 2]
dat_most_recent <- dat %>%
  select(
    "county" = "County Name",
    "state" = "State",
    "most_recent" = "3/27/2020"
  ) %>%
  group_by(state) %>%
  arrange(desc(most_recent)) %>%
  mutate(
    percentage_of_total = round(most_recent / total_cases, digits = 3)
  )

dat_most_recent

```

### Most Recent by County

```

## # A tibble: 3,197 x 4
## # Groups:   state [51]
##   county                state most_recent percentage_of_total
##   <chr>                <chr>      <int>          <dbl>
## 1 Queens County        NY           8214           0.067
## 2 Westchester County   NY           7187           0.059
## 3 Kings County         NY           6750           0.055
## 4 Nassau County        NY           4657           0.038
## 5 Bronx County         NY           4655           0.038
## 6 New York County      NY           4478           0.036
## 7 Suffolk County       NY           3386           0.028
## 8 Cook County          IL           2239           0.018
## 9 Statewide Unallocated NJ           1984           0.016
## 10 King County         WA           1828           0.015
## # ... with 3,187 more rows

```

```

# Get an overview on the data
glimpse_dat <- glimpse(dat)

```

### Raw Data

```

## Observations: 3,197
## Variables: 71
## $ countyFIPS      <int> 0, 1001, 1003, 1005, 1007, 1009, 1011, 1013, 1015, 10...
## $ `County Name`  <chr> "Statewide Unallocated", "Autauga County", "Baldwin C...
## $ State          <chr> "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL", "AL",...
## $ stateFIPS      <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,...
## $ `1/22/2020`    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `1/23/2020`    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...

```

[illegible]

```
## $ `3/18/2020` <int> 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `3/19/2020` <int> 0, 0, 1, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `3/20/2020` <int> 0, 0, 2, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `3/21/2020` <int> 0, 0, 2, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `3/22/2020` <int> 0, 0, 3, 0, 0, 0, 0, 0, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `3/23/2020` <int> 0, 0, 3, 0, 0, 0, 0, 0, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `3/24/2020` <int> 0, 1, 4, 0, 0, 0, 0, 0, 2, 5, 0, 0, 0, 0, 0, 0, 0, 0,...
## $ `3/25/2020` <int> 0, 4, 4, 0, 0, 1, 0, 1, 2, 10, 1, 1, 0, 0, 1, 1, 0, 1...
## $ `3/26/2020` <int> 0, 6, 5, 0, 0, 2, 2, 1, 2, 13, 1, 4, 1, 0, 1, 1, 0, 1...
## $ `3/27/2020` <int> 0, 6, 5, 0, 0, 5, 2, 1, 3, 15, 1, 7, 1, 0, 1, 3, 0, 1...
## $ `3/28/2020` <int> 0, 6, 10, 0, 0, 5, 3, 1, 3, 17, 1, 7, 1, 0, 2, 4, 0, ...
```

glimpse\_dat

```
##      countyFIPS      County Name State stateFIPS 1/22/2020 1/23/2020
## 1:      0 Statewide Unallocated  AL      1      0      0
## 2:     1001 Autauga County  AL      1      0      0
## 3:     1003 Baldwin County  AL      1      0      0
## 4:     1005 Barbour County  AL      1      0      0
## 5:     1007 Bibb County  AL      1      0      0
## ---
## 3193:    56037 Sweetwater County  WY      56      0      0
## 3194:    56039 Teton County  WY      56      0      0
## 3195:    56041 Uinta County  WY      56      0      0
## 3196:    56043 Washakie County  WY      56      0      0
## 3197:    56045 Weston County  WY      56      0      0
##      1/24/2020 1/25/2020 1/26/2020 1/27/2020 1/28/2020 1/29/2020 1/30/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0
## ---
## 3193:      0      0      0      0      0      0      0
## 3194:      0      0      0      0      0      0      0
## 3195:      0      0      0      0      0      0      0
## 3196:      0      0      0      0      0      0      0
## 3197:      0      0      0      0      0      0      0
##      1/31/2020 2/1/2020 2/2/2020 2/3/2020 2/4/2020 2/5/2020 2/6/2020 2/7/2020
## 1:      0      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0      0
## 4:      0      0      0      0      0      0      0      0
## 5:      0      0      0      0      0      0      0      0
## ---
## 3193:      0      0      0      0      0      0      0      0
## 3194:      0      0      0      0      0      0      0      0
## 3195:      0      0      0      0      0      0      0      0
## 3196:      0      0      0      0      0      0      0      0
## 3197:      0      0      0      0      0      0      0      0
##      2/8/2020 2/9/2020 2/10/2020 2/11/2020 2/12/2020 2/13/2020 2/14/2020
## 1:      0      0      0      0      0      0      0
## 2:      0      0      0      0      0      0      0
## 3:      0      0      0      0      0      0      0
```



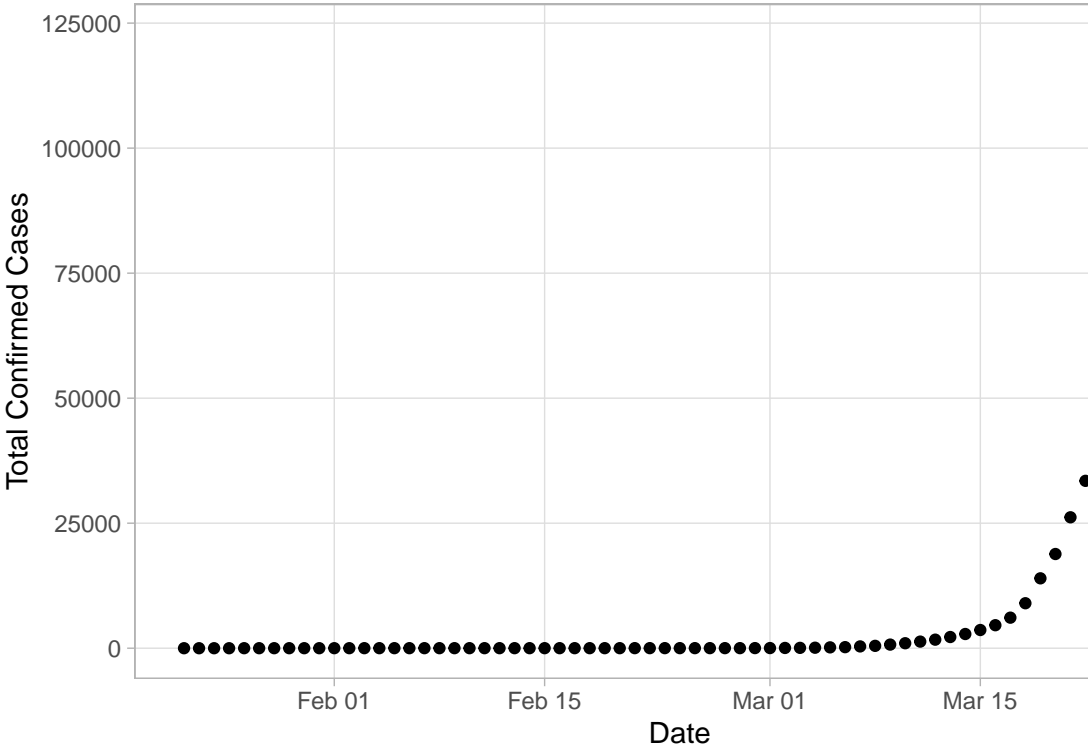
##	4:	0	0	0	0	0	0	0
##	5:	0	0	0	0	0	0	0
##	---							
##	3193:	0	0	0	0	0	0	0
##	3194:	0	0	0	0	0	0	0
##	3195:	0	0	0	0	0	0	0
##	3196:	0	0	0	0	0	0	0
##	3197:	0	0	0	0	0	0	0
##		2/15/2020	2/16/2020	2/17/2020	2/18/2020	2/19/2020	2/20/2020	2/21/2020
##	1:	0	0	0	0	0	0	0
##	2:	0	0	0	0	0	0	0
##	3:	0	0	0	0	0	0	0
##	4:	0	0	0	0	0	0	0
##	5:	0	0	0	0	0	0	0
##	---							
##	3193:	0	0	0	0	0	0	0
##	3194:	0	0	0	0	0	0	0
##	3195:	0	0	0	0	0	0	0
##	3196:	0	0	0	0	0	0	0
##	3197:	0	0	0	0	0	0	0
##		2/22/2020	2/23/2020	2/24/2020	2/25/2020	2/26/2020	2/27/2020	2/28/2020
##	1:	0	0	0	0	0	0	0
##	2:	0	0	0	0	0	0	0
##	3:	0	0	0	0	0	0	0
##	4:	0	0	0	0	0	0	0
##	5:	0	0	0	0	0	0	0
##	---							
##	3193:	0	0	0	0	0	0	0
##	3194:	0	0	0	0	0	0	0
##	3195:	0	0	0	0	0	0	0
##	3196:	0	0	0	0	0	0	0
##	3197:	0	0	0	0	0	0	0
##		2/29/2020	3/1/2020	3/2/2020	3/3/2020	3/4/2020	3/5/2020	3/6/2020
##	1:	0	0	0	0	0	0	0
##	2:	0	0	0	0	0	0	0
##	3:	0	0	0	0	0	0	0
##	4:	0	0	0	0	0	0	0
##	5:	0	0	0	0	0	0	0
##	---							
##	3193:	0	0	0	0	0	0	0
##	3194:	0	0	0	0	0	0	0
##	3195:	0	0	0	0	0	0	0
##	3196:	0	0	0	0	0	0	0
##	3197:	0	0	0	0	0	0	0
##		3/8/2020	3/9/2020	3/10/2020	3/11/2020	3/12/2020	3/13/2020	3/14/2020
##	1:	0	0	0	0	0	0	0
##	2:	0	0	0	0	0	0	0
##	3:	0	0	0	0	0	0	1
##	4:	0	0	0	0	0	0	0
##	5:	0	0	0	0	0	0	0
##	---							
##	3193:	0	0	0	0	0	0	0
##	3194:	0	0	0	0	0	0	0
##	3195:	0	0	0	0	0	0	0

## 3196:	0	0	0	0	0	0	0
## 3197:	0	0	0	0	0	0	0
##	3/15/2020	3/16/2020	3/17/2020	3/18/2020	3/19/2020	3/20/2020	3/21/2020
## 1:	0	0	0	0	0	0	0
## 2:	0	0	0	0	0	0	0
## 3:	1	1	1	1	1	2	2
## 4:	0	0	0	0	0	0	0
## 5:	0	0	0	0	0	0	0
## ---							
## 3193:	0	0	0	0	0	0	0
## 3194:	0	0	0	1	1	2	2
## 3195:	0	0	0	0	0	0	0
## 3196:	0	0	0	0	0	0	0
## 3197:	0	0	0	0	0	0	0
##	3/22/2020	3/23/2020	3/24/2020	3/25/2020	3/26/2020	3/27/2020	3/28/2020
## 1:	0	0	0	0	0	0	0
## 2:	0	0	1	4	6	6	6
## 3:	3	3	4	4	5	5	10
## 4:	0	0	0	0	0	0	0
## 5:	0	0	0	0	0	0	0
## ---							
## 3193:	0	0	1	1	1	1	1
## 3194:	2	2	3	6	8	12	14
## 3195:	0	0	0	0	0	0	0
## 3196:	0	0	0	0	0	1	1
## 3197:	0	0	0	0	0	0	0

## Basic Visualization

```
ggplot(total_confirmed_daily, aes(x = date, y = total_confirmed)) +
  geom_point() +
  theme_light() +
  theme(
    panel.grid.minor = element_blank(),
  ) +
  ggtitle(
    label = "Total Confirmed Cases, Nation-wide, Daily",
    subtitle = "COVID-19 GLM Model Project"
  ) +
  xlab("Date") +
  ylab("Total Confirmed Cases")
```

Total Confirmed Cases, Nation-wide, Daily  
COVID-19 GLM Model Project



Daily Nation-wide Total