PLSC 476: Empirical Legal Studies

Christopher Zorn

September 4, 2025

R and RMarkdown

- RMarkdown is "an authoring framework for data science" that facilitates literate programming and reproduceable research
- Purpose: To allow for the creation of *reproduceable*, *dynamic* documents, presentations, and web pages.
- Similar to Markdown (and other lightweight, literate markup languages)
- Combines text, R code, and R output
- Output formats: PDF, HTML, Word, Shiny (for interactive web apps), others
- These slides were created using RMarkdown; the code for doing so is available on the course github repository

Why Plot?

- Know your data

- Catch mistakes

- Learn something. . .

Running Example: COVID-19 in PA

- $N = 67$ counties
- Current as of July 22, 2025
- Source: PA Department of Health

Variables:

- County Name
- Latitude/Longitude
- FIPS code
- Date
- New Cases
- 7-Day Average of New Cases
- Total/Cumulative Cases
- County Population
- New Case Rate (per 100K population)
- 7-Day Average of New Case Rate
- Cumulative Case Rate (per 100K population)
- New Deaths
- Total/Cumulative Deaths
- Cumulative Death Rate (per 100K population)

Table 1: PA COVID-19 Data, 07/22/2021

	FIPS	Date	County	New.Cases	New.Cases.7dayAvg	Cumulative.Cases
2320	42001	07/22/2021	Adams	5	7.1	9820
3805	42003	07/22/2021	Allegheny	51	44.9	103176
5290	42005	07/22/2021	Armstrong	6	1.4	6062
6775	42007	07/22/2021	Beaver	2	6.6	15737
8260	42009	07/22/2021	Bedford	0	0.6	4783
9745	42011	07/22/2021	Berks	18	16.0	49086
11230	42013	07/22/2021	Blair	2	1.7	13646
12715	42015	07/22/2021	Bradford	2	1.4	6204
14200	42017	07/22/2021	Bucks	37	31.0	61405
15685	42019	07/22/2021	Butler	7	6.4	17850
17170	42021	07/22/2021	Cambria	4	2.1	15069
18655	42023	07/22/2021	Cameron	0	0.4	316

Data Summary

```
      FIPS           Date              County            New.Cases
 Min.   :42001   Length:67          Length:67          Min.   :  0.00
 1st Qu.:42034   Class :character   Class :character   1st Qu.:  1.00
 Median :42067   Mode  :character   Mode  :character   Median :  3.00
 Mean   :42067                                         Mean   :  9.73
 3rd Qu.:42100                                         3rd Qu.:  7.50
 Max.   :42133                                         Max.   :132.00
 New.Cases.7dayAvg Cumulative.Cases   Population       New.Case.Rate
 Min.   : 0.10     Min.   :   316    Min.   :   4447   Min.   : 0.00
 1st Qu.: 1.00     1st Qu.:  3805    1st Qu.:  42025   1st Qu.: 1.70
 Median : 2.10     Median :  8166    Median :  84629   Median : 3.30
 Mean   : 7.36     Mean   : 18409    Mean   : 191074   Mean   : 3.74
 3rd Qu.: 7.05     3rd Qu.: 18489    3rd Qu.: 208270   3rd Qu.: 4.95
 Max.   :80.90     Max.   :157903    Max.   :1584064   Max.   :13.80
 New.Case.Rate.7DayAvg Cumulative.Case.Rate   Longitude        Latitude
 Min.   :0.80          Min.   : 6683         Min.   :-80.4   Min.   :39.9
 1st Qu.:1.90          1st Qu.: 8467         1st Qu.:-79.1   1st Qu.:40.4
 Median :2.80          Median : 9529         Median :-77.4   Median :40.8
 Mean   :3.23          Mean   : 9673         Mean   :-77.6   Mean   :40.8
 3rd Qu.:4.00          3rd Qu.:10627         3rd Qu.:-76.2   3rd Qu.:41.3
 Max.   :9.60          Max.   :19829         Max.   :-75.0   Max.   :42.0
 Georeferenced.Lat...Long  New.Deaths        Total.Deaths    Total.Death.Rate
 Length:67                Min.   :0.0000   Min.   :  10.0    Min.   : 96.8
 Class :character         1st Qu.:0.0000   1st Qu.:  93.5    1st Qu.:187.8
 Mode  :character         Median :0.0000   Median : 190.0    Median :224.9
                          Mean   :0.0597   Mean   : 419.0    Mean   :229.5
                          3rd Qu.:0.0000   3rd Qu.: 438.0    3rd Qu.:262.8
                          Max.   :1.0000   Max.   :3811.0    Max.   :408.4
```

Univariate Graphics

- Dotcharts / Barcharts
- Histograms
- Density Plots
- Boxplots
- Q-Q Plots
- Others. . .

A Dotchart

Figure 1: New COVID-19 Cases by County

A Sorted Dotchart

Figure 2: New COVID-19 Cases by County
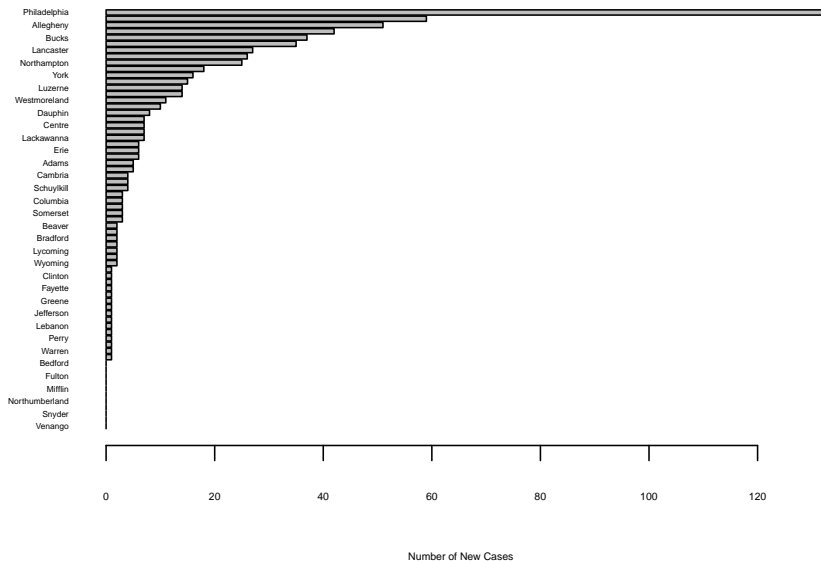
A (Sorted) Barchart

Figure 3: New COVID-19 Cases by County

The Histogram: Cumulative Case Percentages

```
   vars  n mean   sd median trimmed  mad  min  max range skew kurtosis   se
X1    1 67 9.67 1.91   9.53    9.54 1.63 6.68 19.8  13.2 2.19     9.79 0.23
```
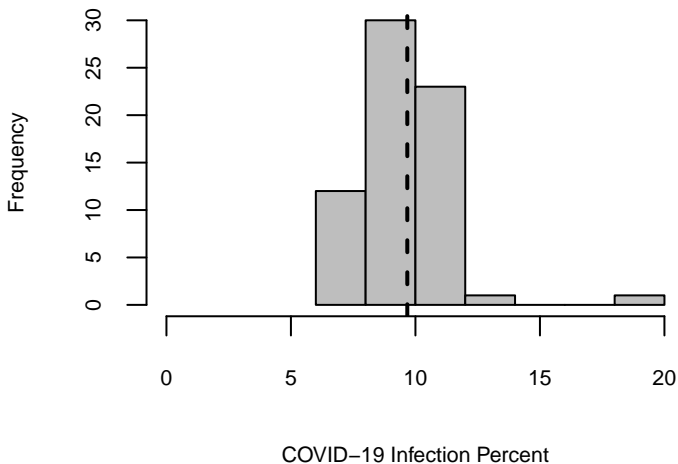

COVID−19 Infection Percent

Figure 4: Cumulative Cases, as a Percentage of Population

"Kernel Density" Plot

COVID−19 Infection Percent
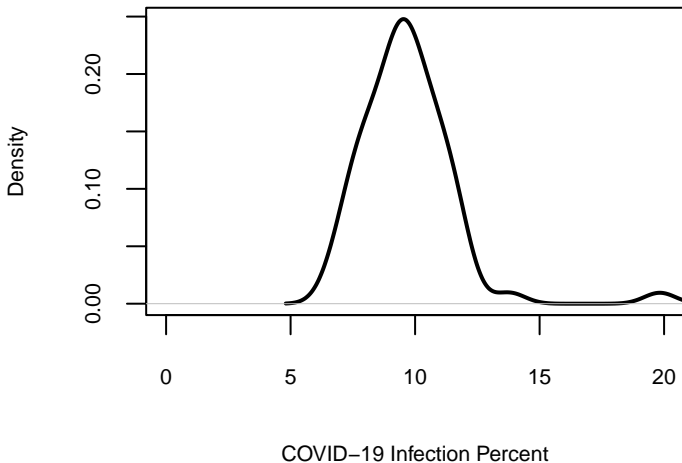
Figure 5: Cumulative Cases, as a Percentage of Population

Density + Histogram

COVID−19 Infection Percent
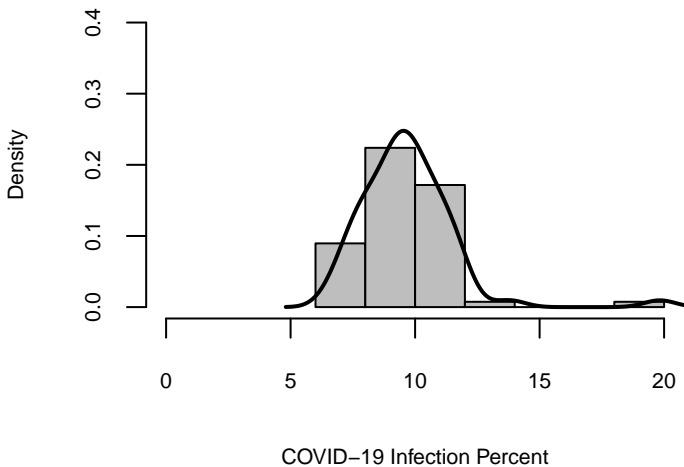
Figure 6: Cumulative Cases, as a Percentage of Population

A Boxplot
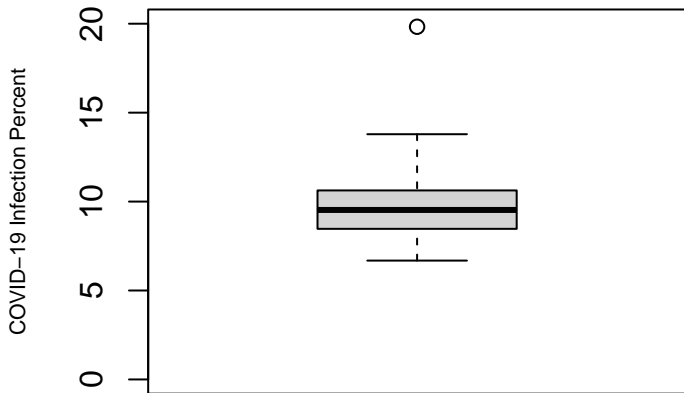
Figure 7: Cumulative Cases, as a Percentage of Population

Multiple Boxplots in One Figure
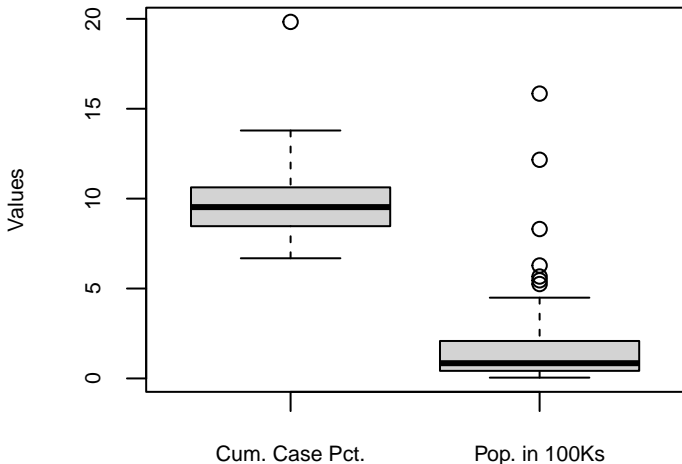
Figure 8: Two Boxplots

Quantile-Quantile (Q-Q) Plots

- Plots two sets of quantiles against each other...
- Can be used to compare two variables' distributions to each other
- Can also compare the *empirical* distribution of a variable to a *theoretical* distribution
- If the two are the same, the quantiles will lie on a straight line

One Density. . .

New COVID Cases per 100K
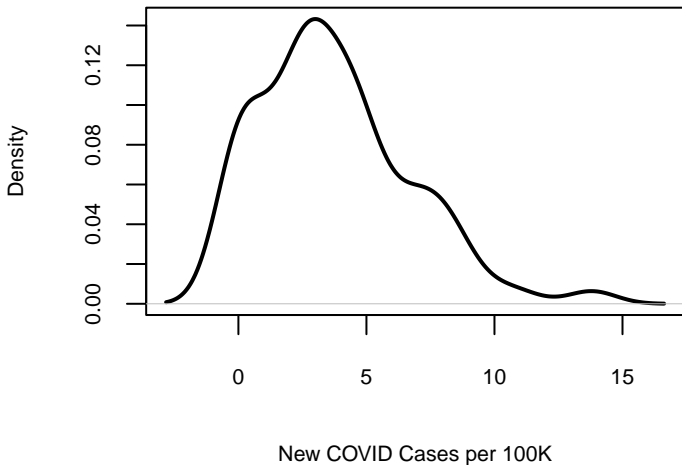
Figure 9: New Cases per 100K Population

One-Variable Normal Q-Q Plot

Theoretical Quantiles
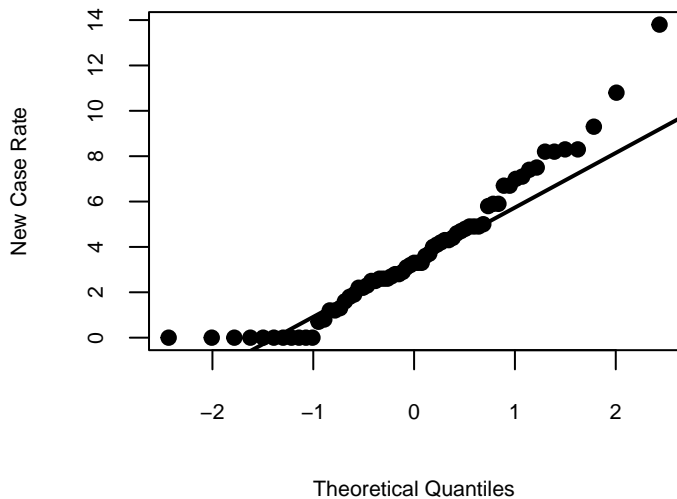
Figure 10: New Cases per 100K Population

Two Densities. . .
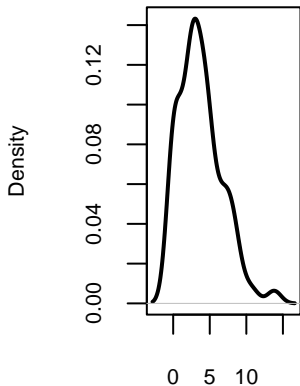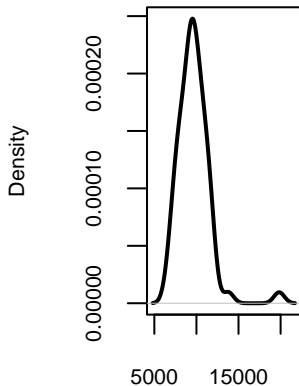
New COVID Cases per 100K   Cumulative COVID Cases per 100

Figure 11: New and Cumulative Cases per 100K Population

Two-Variable Q-Q Plot
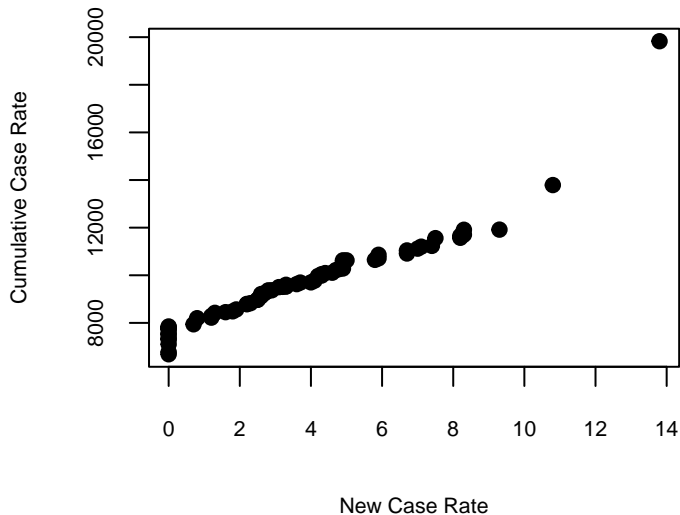
Figure 12: New and Cumulative Cases per 100K Population

Other Univariate Plots

- Pie charts (please don't. . .)
- "Donut" plots (same)
- "Stem-and-leaf" plots (very old-school)
- Stripplots
- *Time-Series Plots*. . .

A Time Series Plot

Date
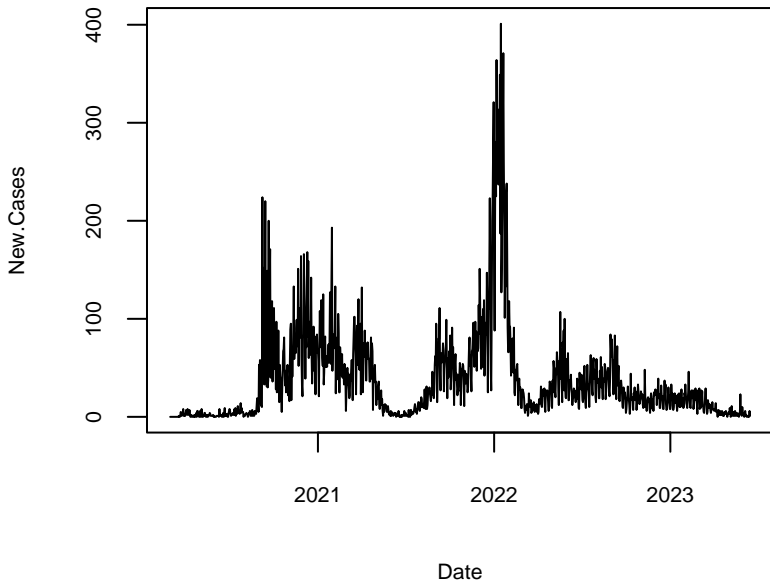
Figure 13: Daily COVID Cases in Centre County, 3/1/2020-6/15/2023

Add a "Smoother"

Date
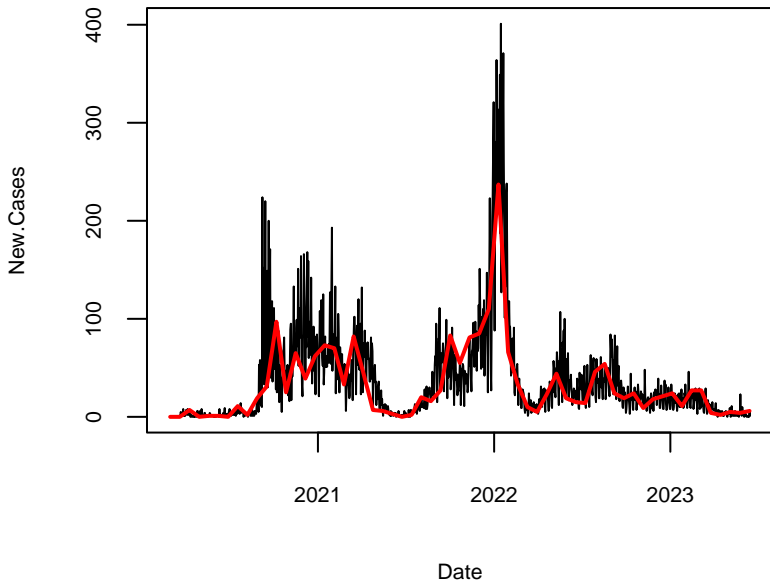
Figure 14: Daily COVID Cases in Centre County, 3/1/2020-6/15/2023

Another Cool Viz: Maps
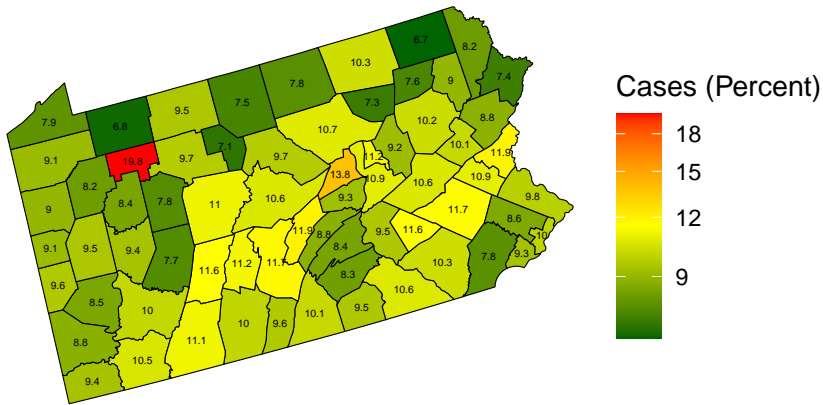
Figure 15: Total Cases (Percent of the Population)

Bivariate Plots: The Scatterplot

Cumulative.Case.Rate
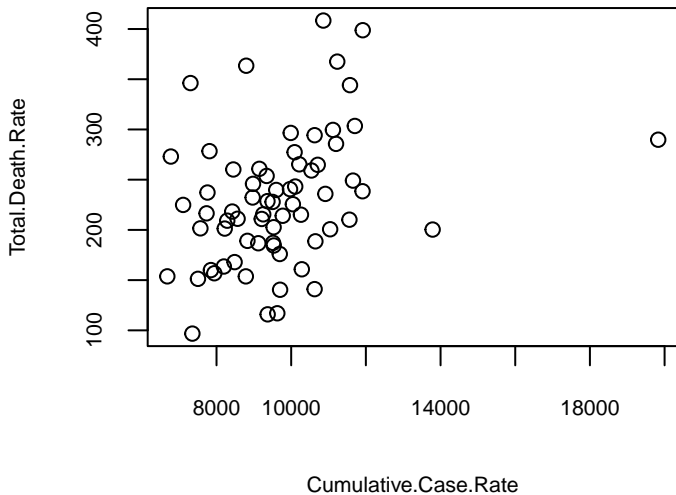
Figure 16: Case Rates vs. Death Rates

A Better Scatterplot
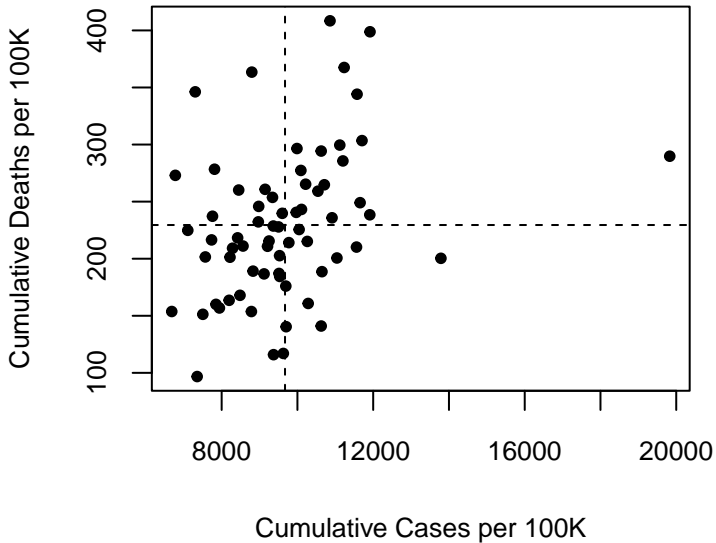
Figure 17: Case Rates vs. Death Rates

Rescaling Axes (Log Scales)
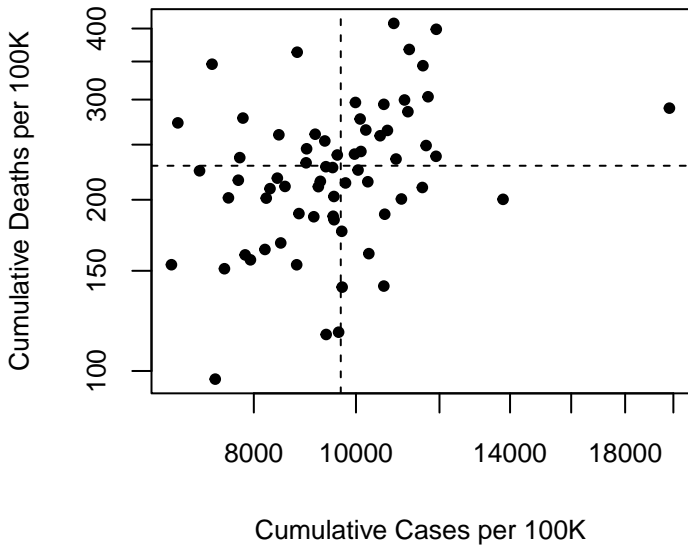
Figure 18: Case Rates vs. Death Rates

Adding Lines
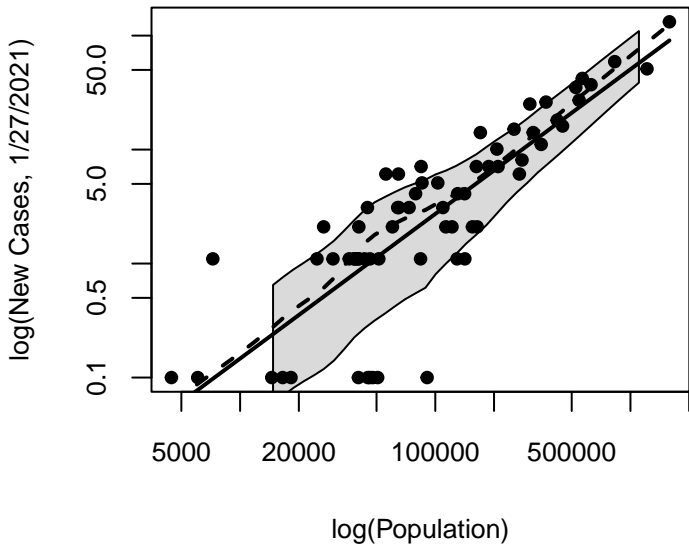
Figure 19: New Cases vs. Population

How Not To Draw A Scatterplot
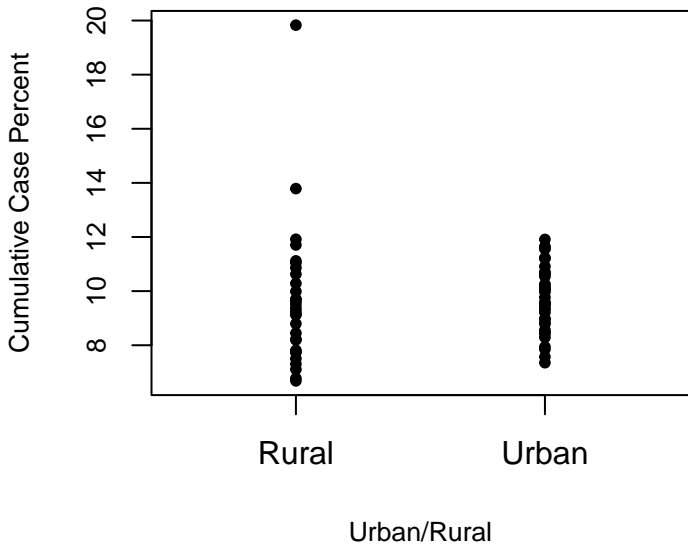
Urban/Rural
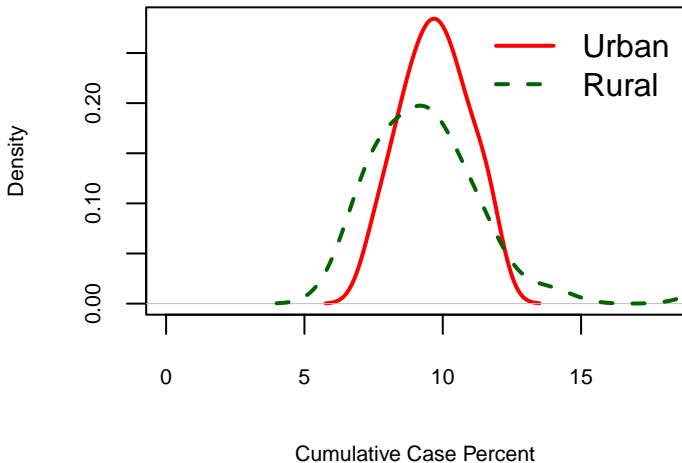
Figure 20: Urban/Rural vs. Cumulative Case Percent

Better. . .

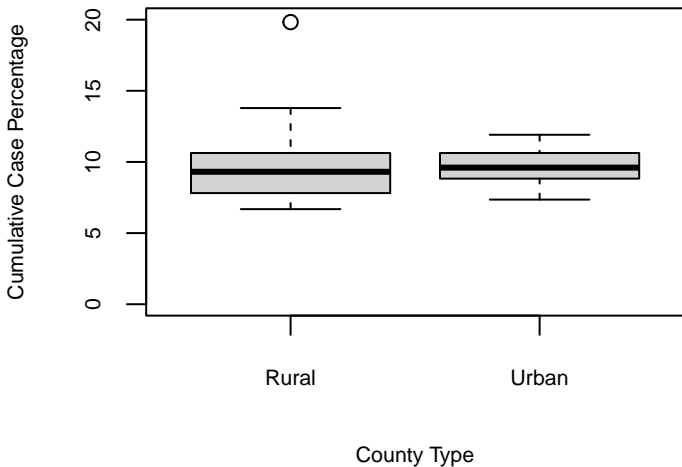Figure 21: Cumulative Cases: Percent of the Population

Or:

County Type

Figure 22: Cumulative Cases: Percent of the Population
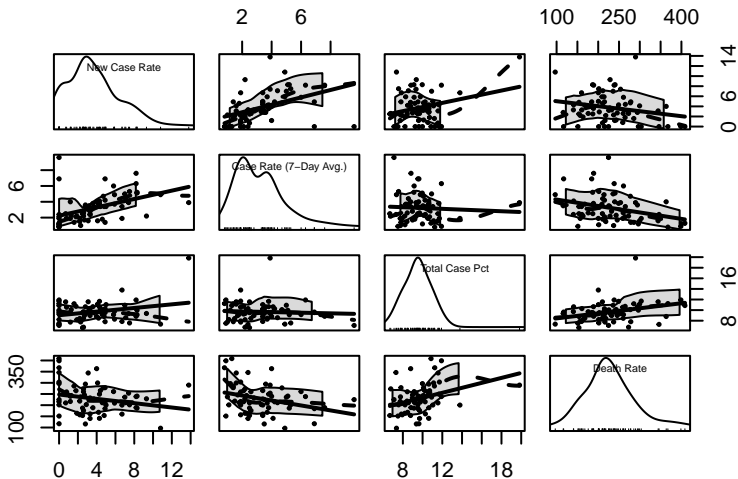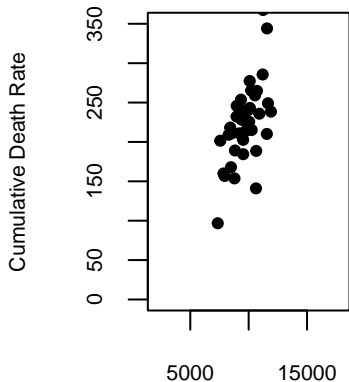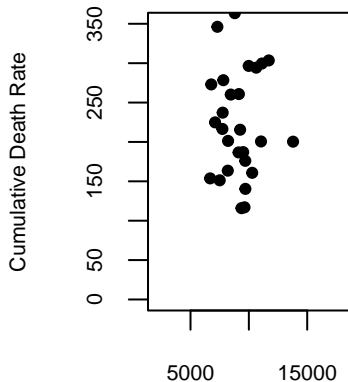
Multivariate Plots: Scatterplot Matrix

Figure 23: COVID in Pennsylvania (1/27/2021)

Conditioned Scatterplots

Figure 24: Case and Death Rates, by Urban/Rural

Other Cool Visualizations

- Contour / Wireframe Plots ("3D")
- Radar plots
- Parallel coordinates plots
- Dynamic / interactive graphics
- Many more. . .

Good Visualization Practices

- Go from simple to complex
- Optimize {readability, information density}
- Each plot should "stand on its own"
- Consider colorblindness (see, e.g., here)