

PLSC476: Empirical Legal Studies

Christopher Zorn

January 28, 2021

R and RMarkdown

- [RMarkdown](#) is “an authoring framework for data science” that facilitates literate programming and reproducible research
- Purpose: To allow for the creation of *reproducible, 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 yesterday (1/27/2021)
- Source: [PA Department of Health](#)

Variables:

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

Table 1: PA COVID-19 Data, 1/27/2021

	FIPS	Date	County	New.Cases	New.Cases.7dayAvg	Cumulative.Cases
362	42001	01/27/2021	Adams	50	50.7	6451
697	42003	01/27/2021	Allegheny	358	345.0	68451
1032	42005	01/27/2021	Armstrong	23	21.0	4557
1367	42007	01/27/2021	Beaver	67	51.7	10830
1702	42009	01/27/2021	Bedford	13	18.7	3587
2037	42011	01/27/2021	Berks	742	277.1	31367
2372	42013	01/27/2021	Blair	31	41.0	9810
2707	42015	01/27/2021	Bradford	34	25.7	3866
3042	42017	01/27/2021	Bucks	286	243.1	38989
3377	42019	01/27/2021	Butler	79	74.3	12568
3712	42021	01/27/2021	Cambria	32	45.0	10666
4047	42023	01/27/2021	Cameron	0	2.3	237

Data Summary

FIPS	Date	County	New.Cases
Min. :42001	Length:67	Length:67	Min. : 0
1st Qu.:42034	Class :character	Class :character	1st Qu.: 14
Median :42067	Mode :character	Mode :character	Median : 38
Mean :42067			Mean :104
3rd Qu.:42100			3rd Qu.: 78
Max. :42133			Max. :742
New.Cases.7dayAvg	Cumulative.Cases	Population	New.Case.Rate
Min. : 1	Min. : 237	Min. : 4447	Min. : 0.0
1st Qu.: 14	1st Qu.: 2476	1st Qu.: 42025	1st Qu.: 27.9
Median : 37	Median : 5845	Median : 84629	Median : 36.7
Mean : 77	Mean : 12323	Mean : 191074	Mean : 48.7
3rd Qu.: 86	3rd Qu.: 12459	3rd Qu.: 208270	3rd Qu.: 47.9
Max. :437	Max. :103022	Max. :1584064	Max. :224.8
New.Case.Rate.7DayAvg	Cumulative.Case.Rate	Longitude	Latitude
Min. : 11.8	Min. : 3680	Min. : -80.4	Min. : 39.9
1st Qu.: 31.9	1st Qu.: 5693	1st Qu.: -79.1	1st Qu.: 40.4
Median : 38.4	Median : 6606	Median : -77.4	Median : 40.8
Mean : 41.7	Mean : 6661	Mean : -77.6	Mean : 40.8
3rd Qu.: 51.4	3rd Qu.: 7192	3rd Qu.: -76.2	3rd Qu.: 41.3
Max. :112.5	Max. :17262	Max. : -75.0	Max. :42.0
Georeferenced.Lat...Long	New.Deaths	Total.Deaths	Total.Death.Rate
Length:67	Min. :0.00	Min. : 7	Min. : 79
Class :character	1st Qu.:0.00	1st Qu.: 72	1st Qu.:140
Mode :character	Median :0.00	Median : 154	Median :172
	Mean :0.39	Mean : 322	Mean :175
	3rd Qu.:1.00	3rd Qu.: 347	3rd Qu.:195
	Max. :4.00	Max. :2905	Max. :334

Univariate Graphics

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

A Dotchart

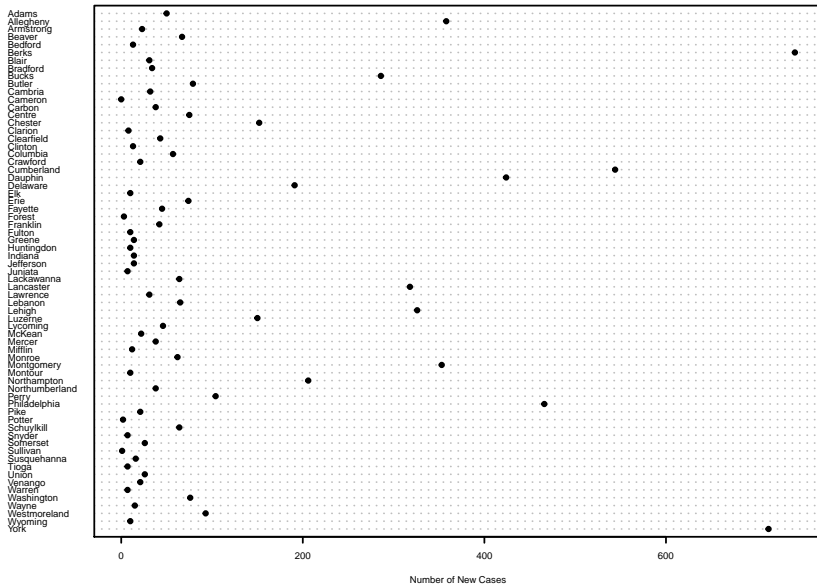


Figure 1: New COVID-19 Cases by County

A Sorted Dotchart

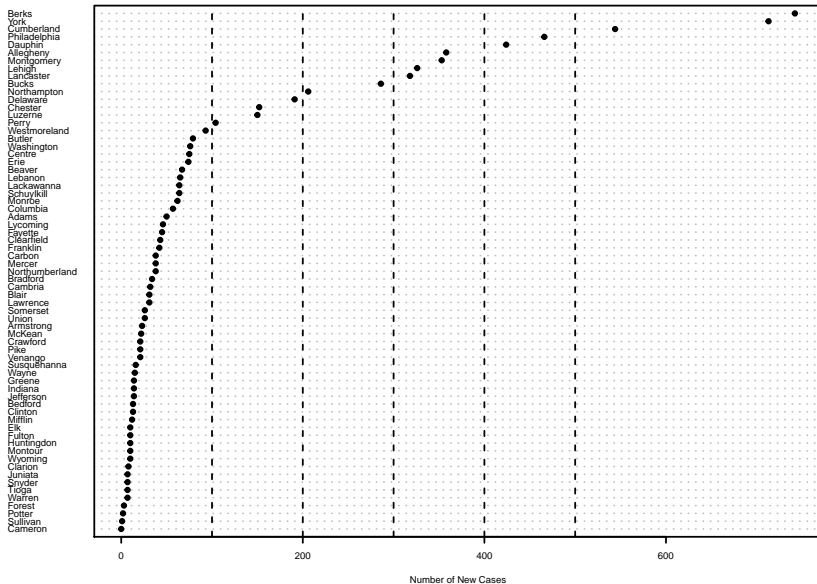


Figure 2: New COVID-19 Cases by County

A (Sorted) Barchart

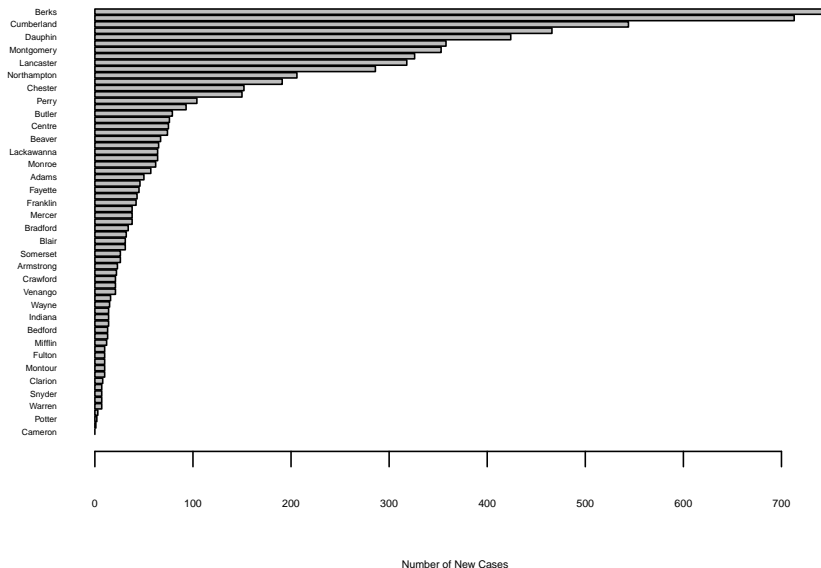


Figure 3: New COVID-19 Cases by County

The Histogram: Cumulative Case Percentages

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
X1	1	67	6.66	1.76	6.61	6.55	1.17	3.68	17.3	13.6	3.08	17.3	0.22

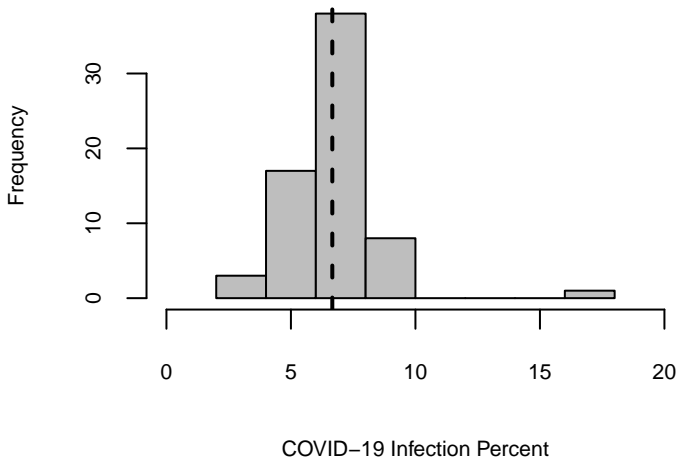


Figure 4: Cumulative Cases, as a Percentage of Population

“Kernel Density” Plot

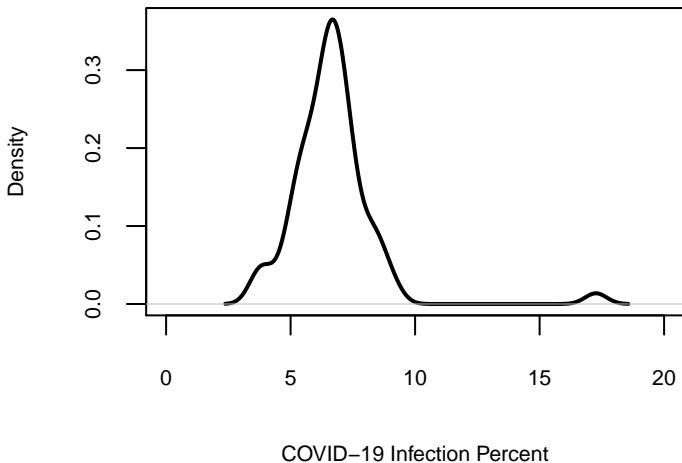


Figure 5: Cumulative Cases, as a Percentage of Population

Density + Histogram

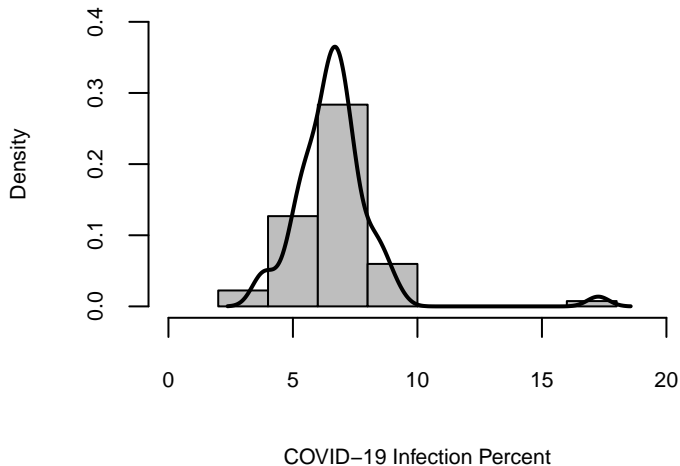


Figure 6: Cumulative Cases, as a Percentage of Population

A Boxplot

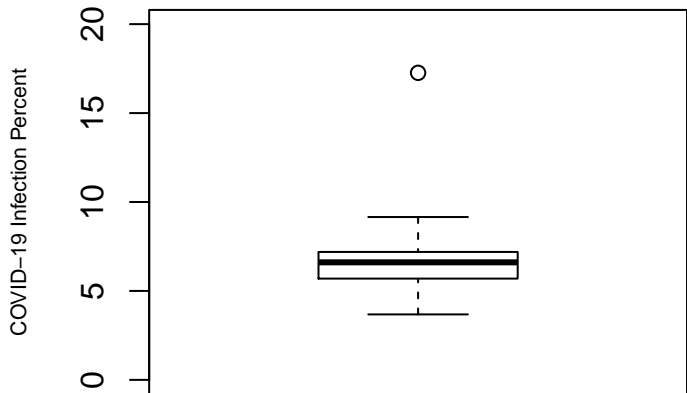


Figure 7: Cumulative Cases, as a Percentage of Population

Multiple Boxplots in One Figure

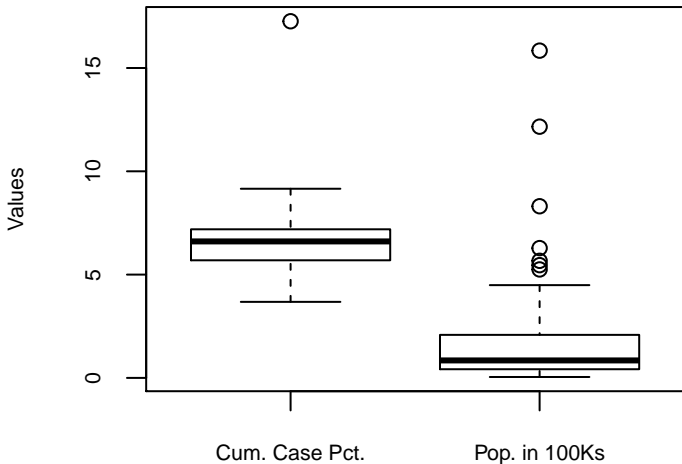


Figure 8: Two Boxplots

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...

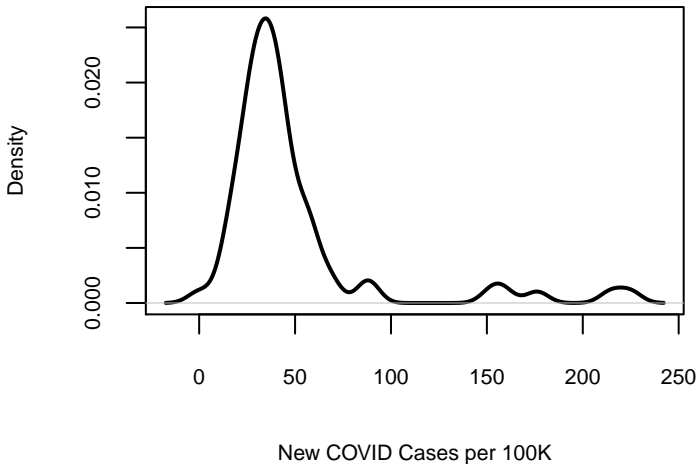


Figure 9: New Cases per 100K Population

One-Variable Normal Q-Q Plot

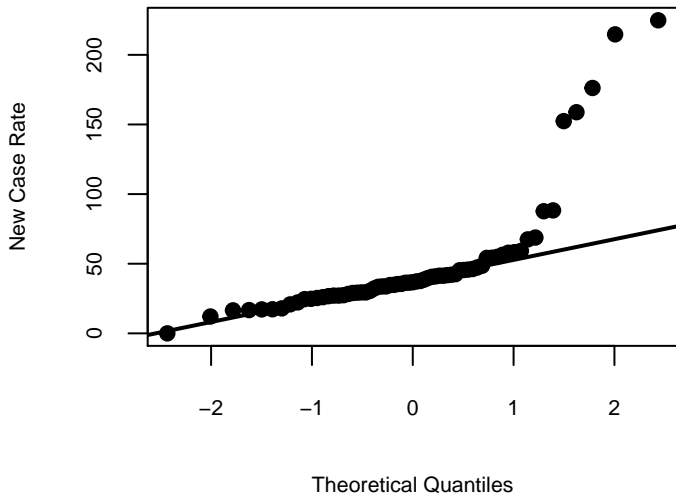
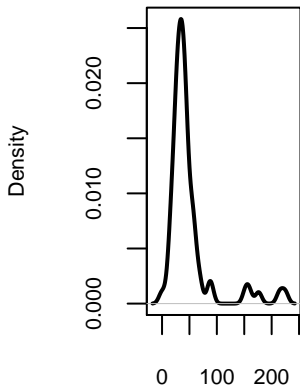
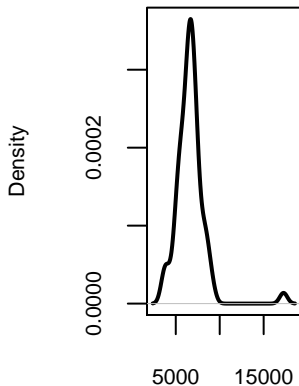


Figure 10: New Cases per 100K Population

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

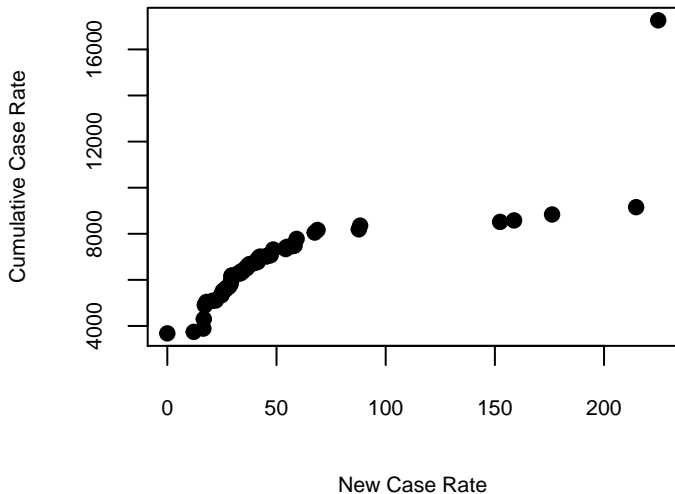


Figure 12: New and Cumulative Cases per 100K Population

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

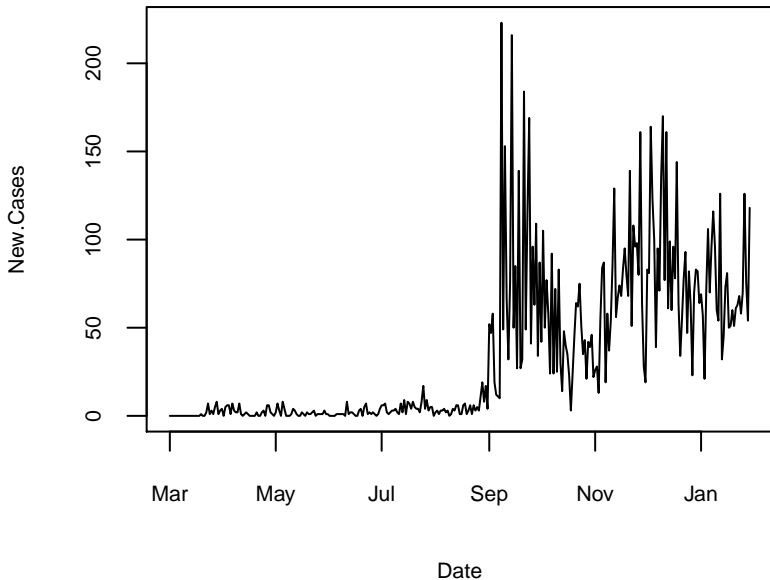


Figure 13: Daily COVID Cases in Centre County, 3/1/2020-1/27/2021

Add a “Smoother”

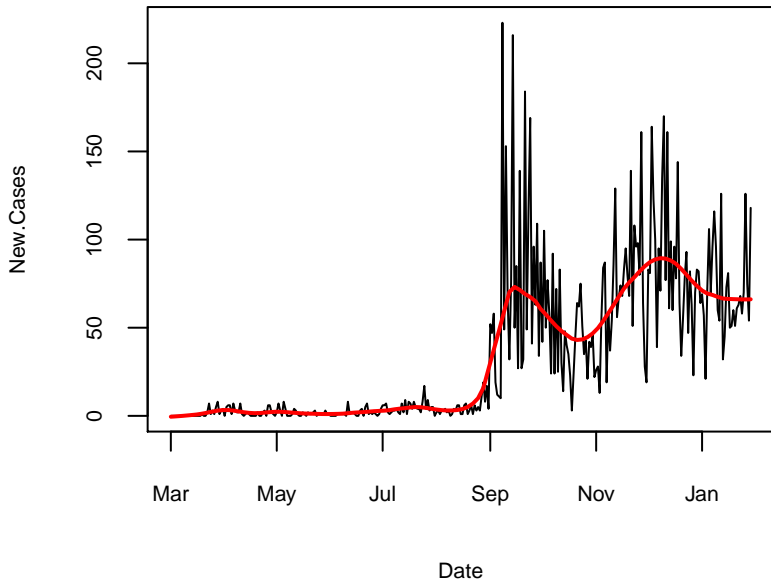


Figure 14: Daily COVID Cases in Centre County, 3/1/2020-1/27/2021

Another Cool Viz: Maps

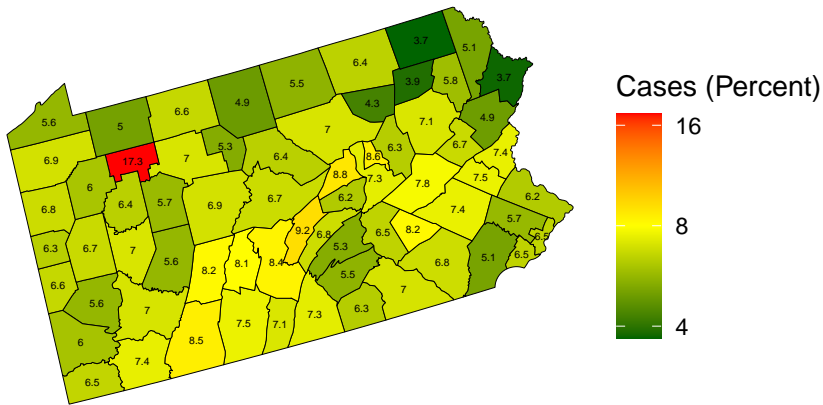


Figure 15: Total Cases (Percent of the Population)

Bivariate Plots: The Scatterplot

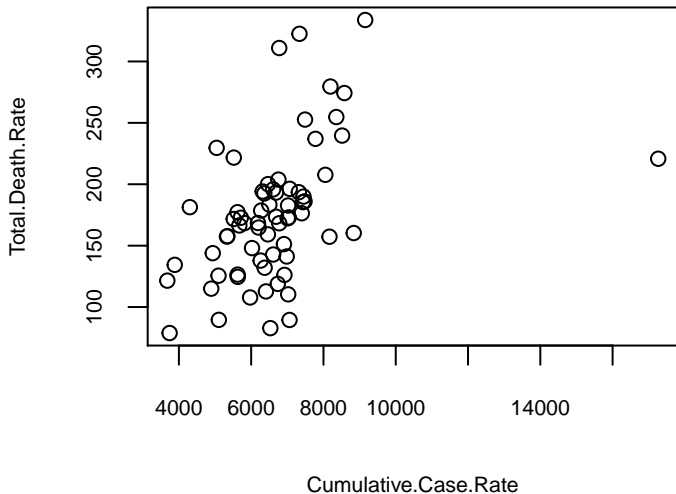


Figure 16: Case Rates vs. Death Rates

A Better Scatterplot

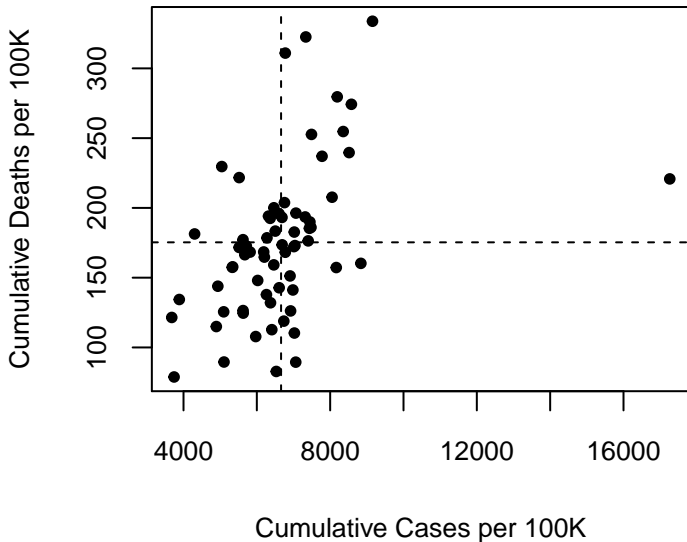


Figure 17: Case Rates vs. Death Rates

Rescaling Axes (Log Scales)

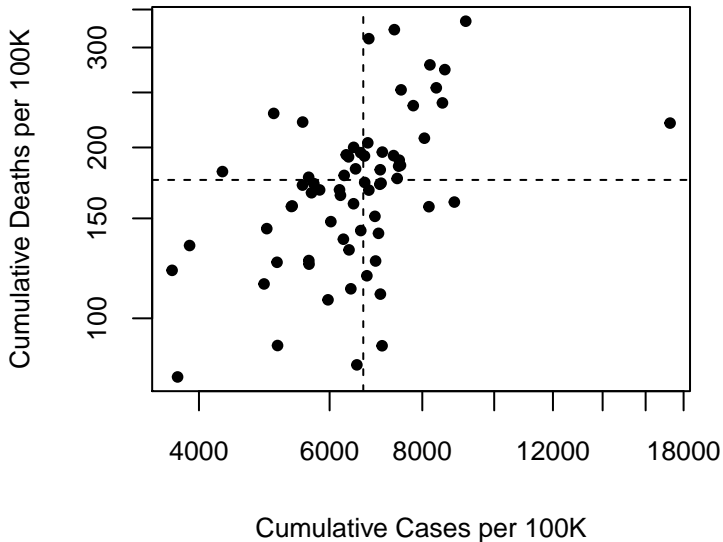


Figure 18: Case Rates vs. Death Rates

Adding Lines

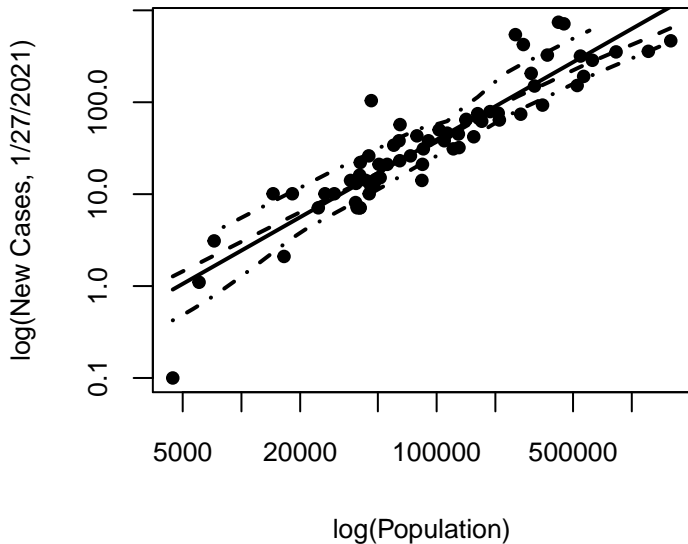


Figure 19: New Cases vs. Population

How Not To Draw A Scatterplot

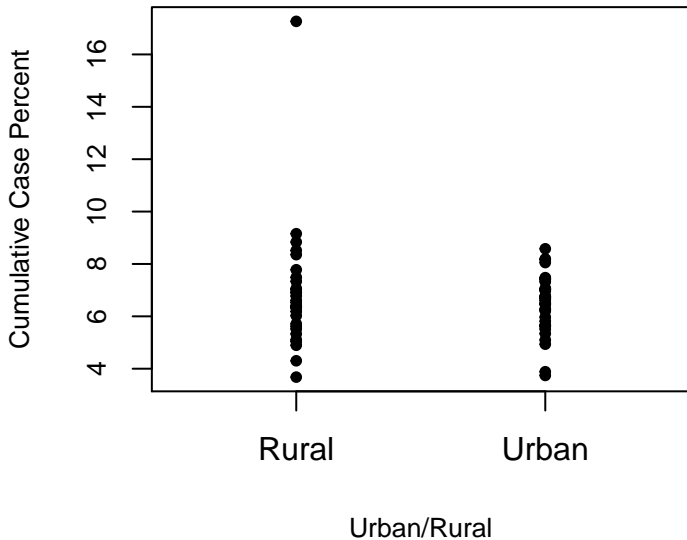


Figure 20: Urban/Rural vs. Cumulative Case Percent

Better...

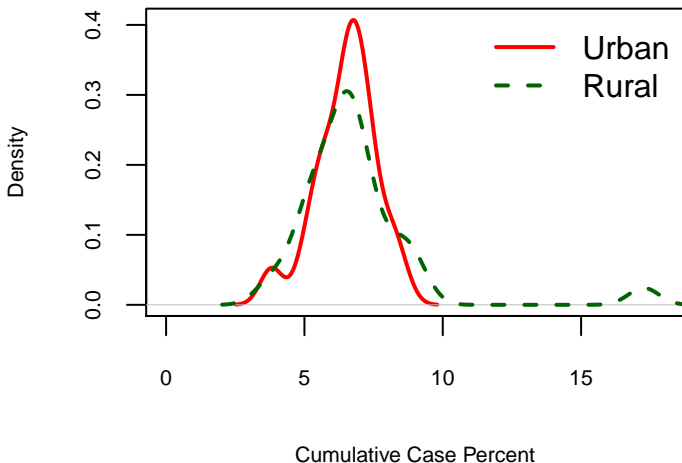


Figure 21: Cumulative Cases: Percent of the Population

Or:

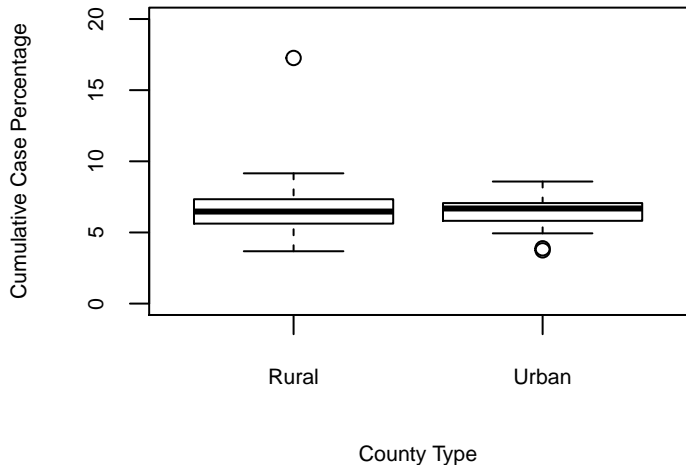


Figure 22: Cumulative Cases: Percent of the Population

Multivariate Plots: Scatterplot Matrix

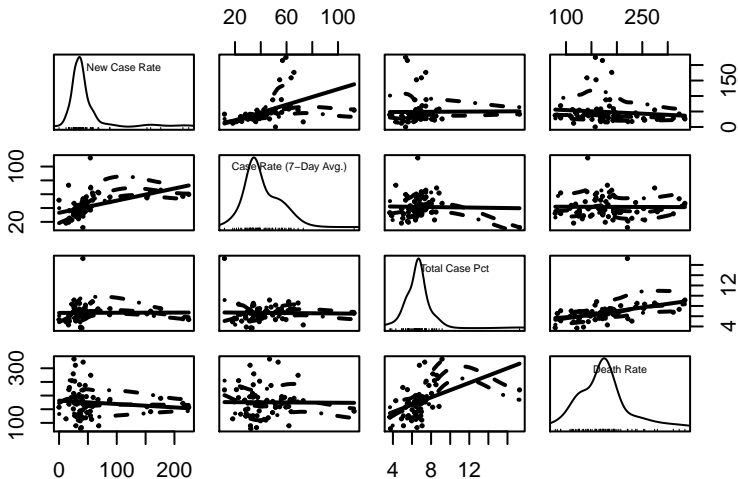
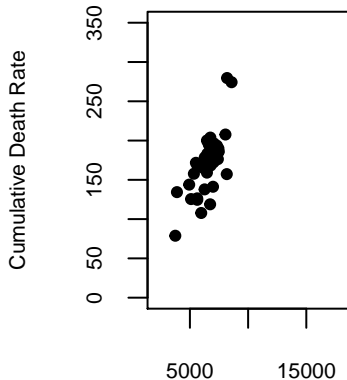
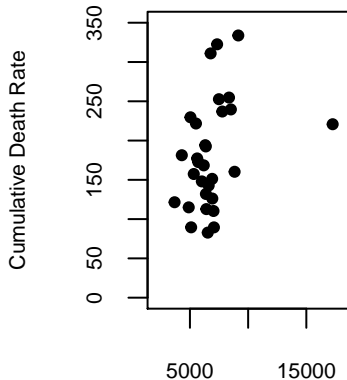


Figure 23: COVID in Pennsylvania (1/27/2021)

Conditioned Scatterplots



Cumulative Case Rate: Urban



Cumulative Case Rate: Rural

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. . .