Legionellosis: County and Population in Ohio

David A. Hibler

December 2, 2019

Contents

1 Legionellosis

Legionellosis is a set of two respiratory diseases outcomes both caused by infections from the bactria *Legionella*; Legionnaires' disease (LD), which is a community pneumonia and has a mortality rate of up to 30 percent (as high as 80 percent, with extreme cases), and Pontiac Fever, a non-pneumonia, flu-like illness that often self-limits after two to five days.

Legionella infections are rising dramatically, with *Legionella pneumophila* (*L. pneumophila*) representing a large proportion of the more than 40,000 water-borne disease cases reported annually. 8,000 to 18,000 hospitalizations of suspected *L. pneumophila* occur every year in the US, at an expense of more than 433.8 million dollars in healthcare costs alone. Rural water distribution systems with older water age, compared to urban ones, often have more favorable conditions for the reproduction of a wide variety of microorganisms, including opportunistic pathogens such as *L. pneumophila*. However, concentrations of *L. pneumophila* and rates of its associated diseases are presenting with much higher rates in urban areas as opposed to rural areas.

Reasons for this are unknown, but this analysis examines the trends of legionellosis infections in Ohio counties by population.

2 Data

The data used in this analysis was collected from several online data bases:

- US Census Bureau
- World Population Review
- Ohio Department of Health

2.1 Data Processing

- pop_2010: Ohio county population from 2010 census.
- pop_2018: Estimated Ohio county population. Retrieved 2019-01-24, from worldpopulationreview.com
- pop_dens: Population density calculated from 2018 population and county area in sq miles.
- ln_2018_pop: The calculated natural log of the 2018 population.
- ln_pop_density: The calculated natural log of the 2018 population density.
- pop_dif: The calculated change in population from 2010 to 2018.
- abs_pop_dif: The absolute value of change in population from 2010 to 2018.
- l_2017: Reported Legionellosis infections in 2017 gathered from the Ohio Department of Health Infections Diseases Report.
- l_2018: Reported Legionellosis infections in 2018 that was specially disclosed by the Ohio Department of Health and is not yet available online, so was not used in this project.

Data set calculations (such as population densities and natural logs) were done in excel and then the data set was saved as a csv file to be analyzed in R.

3 Analysis

Data sets were plotted for visual examination, and correlational tests (such as Pearson's, Kendall's, and Spearman's) were run to check for significance.

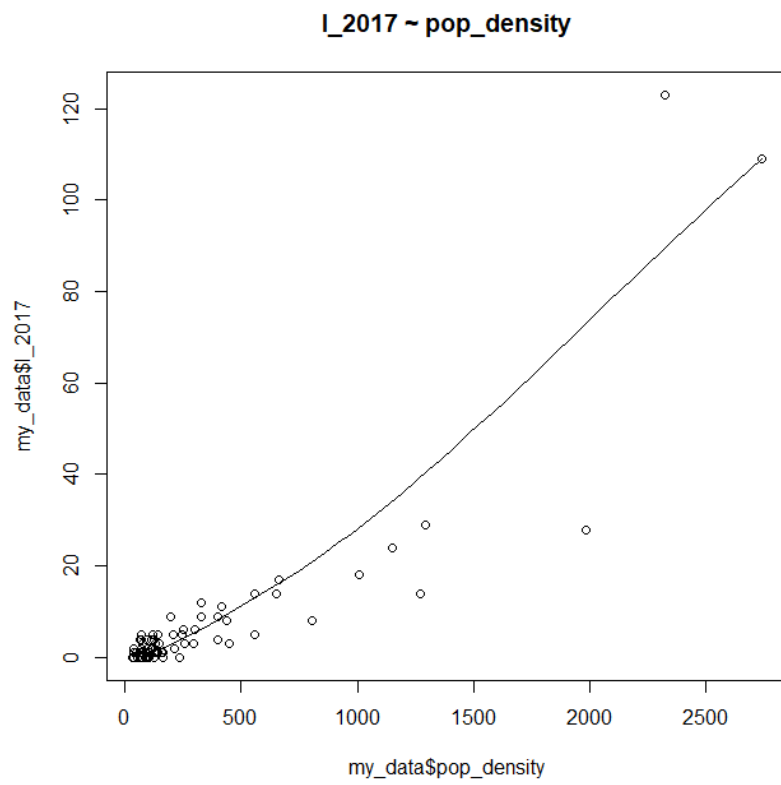| test	Infect vs Population	Infect vs Pop_density	Infect vs Ln_Pop	Infect vs Ln_Pop_density
Pearson	1.43E-39	2.11E-30	5.44E-12	1.10E-11
Kendall	4.89E-14	1.33E-13	4.89E-14	1.33E-13
Spearman	1.73E-15	1.07E-14	1.73E-15	1.07E-14

3.1 Plots
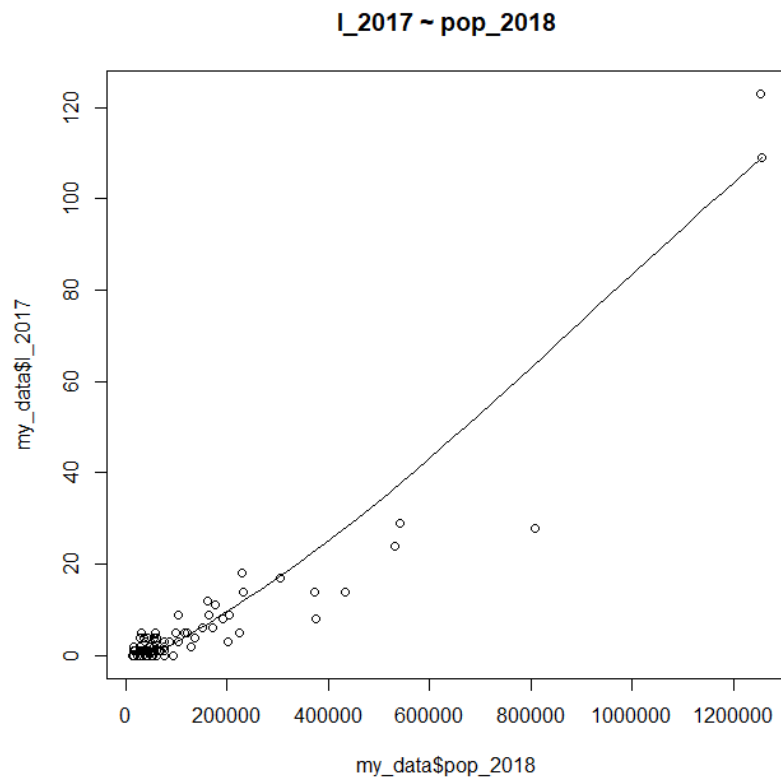
Figure 1: Infections vs population

I_2017 ~ pop_density

Figure 2: Infections vs population density

I_2017 ~ ln_2018_pop
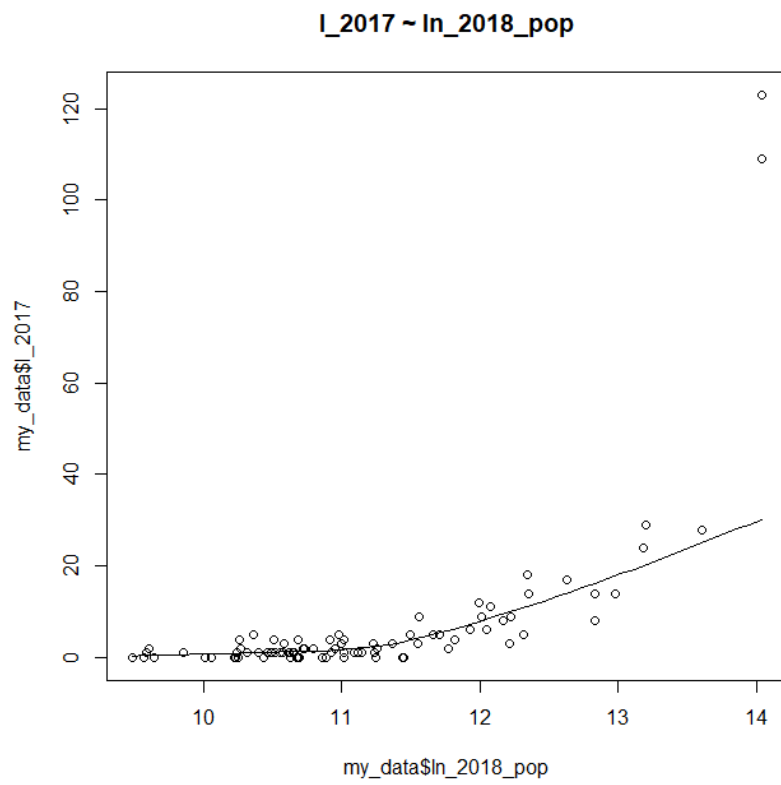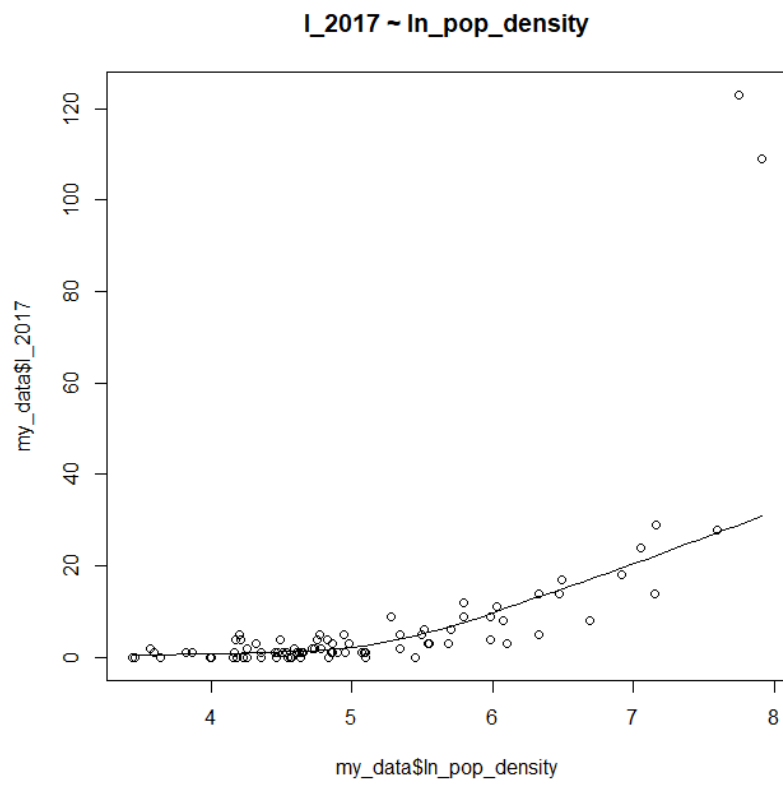
Figure 3: Infections vs Natural log of population

Figure 4: Infections vs Natural log of population density

4 Conclusion

Rates of Legionellosis were very significantly related to population size, density and the natural logs of both. Population size did seem to have the highest degree of significance. So this does provide evidence that this non-communicable disease is a greater public health issue in high population environments (urban), than it is in less populated environments (rural). However, this data does not provide much more insight beyond that.

The trends on the scatter plots do look to be non-linear, possibly suggesting that more heavily populated areas add compounding effects to infection rates. However additional research and analysis would be needed to determine what effects are impacting these results.