Cancer EDA

Vinicio De Sola, Sam Tosaria, Joanna Yu

September 26, 2018

Introduction

Question

- Help governments allocate resources in fighting Cancer
- Variables that could be related: Socio economic and geographic / environmental
- Data process: County level, overrepresentation?

Data

No.	Category	Fields
1	Income	MedIncome, povertyPercent, binnedInc
2	Population, Age, Gender	popEst2015, MedianAge, MedianAgeMale, MedianAgeFemale
3	Household Related	AvgHouseholdSize, PercentMarried, PctMarriedHouseholds
4	Education and Age	PctNoHS18_24, PctHS18_24, PctSomeCol18_24, PctBachDeg18_24, PctHS25_Over, PctBachDeg25_Over
5	Employment Status	PctEmployed16_Over, PctUnemployed16_Over
6	Insurance	PctPrivateCoverage, PctEmpPrivCoverage, PctPublicCoverage

Cleaning Process

- Creation of new field 'Cancer per 1000'
- Change the Mortality rate to per 1000
- Average Household Size, Median Age, and Cancer Occurrence had issues
- Data lost: 9%. We can live with that!
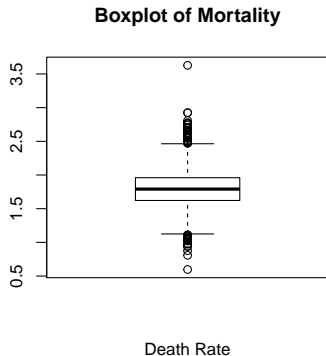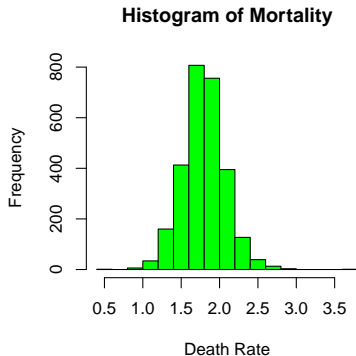
Key Univariate Descriptions

Mortality

Histogram of Mortality

Boxplot of Mortality

Table 2: Higher Outlier

	County	State
1490	Union County	Florida

Median Income

Histogram of Median Income

Boxplot of Median Income
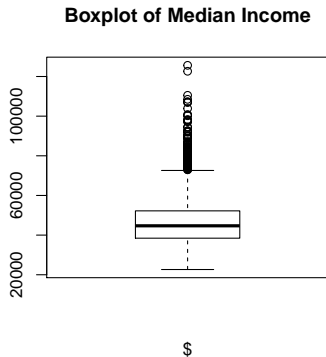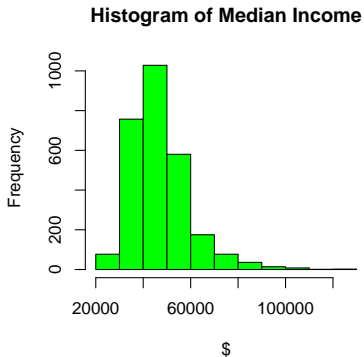
Table 3: Higher Outlier

	County	State
260	Falls Church city	Virginia

Household Size

Histogram of Average Household Size
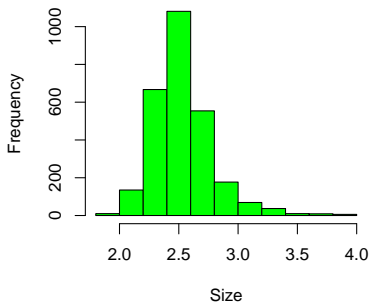
Boxplot of Average Household Size

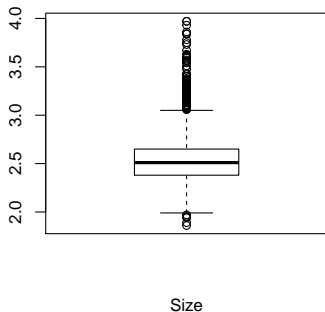Table 4: Higher Outlier

	County	State
176	McKinley County	New Mexico
2729	Northwest Arctic Borough	Alaska

Poverty Line

Table 5: Poverty Line

	Mean	Median	Sd	Min	Max
Poverty Percent	17.2	16.3	6.49	3.2	47.4

Table 6: Poorest County

	County	State
2826	Todd County	South Dakota

Key Relationships

Geography

Top 15 States by mortality
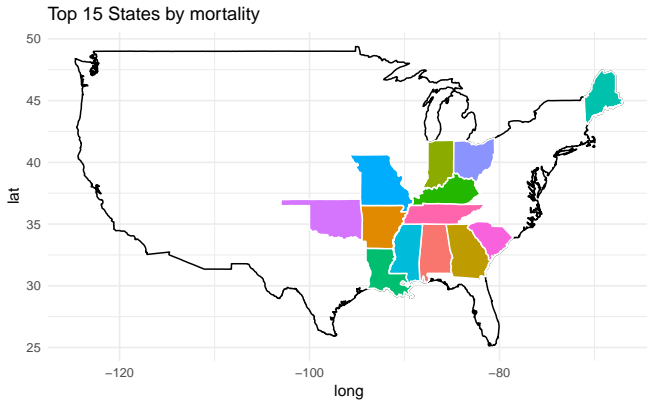
- ▶ Interesting Result!

Geography

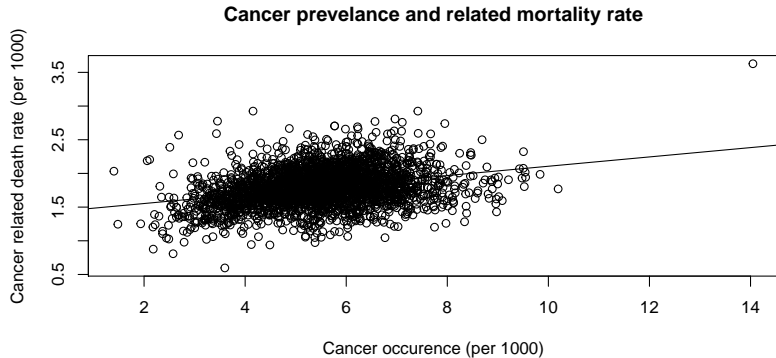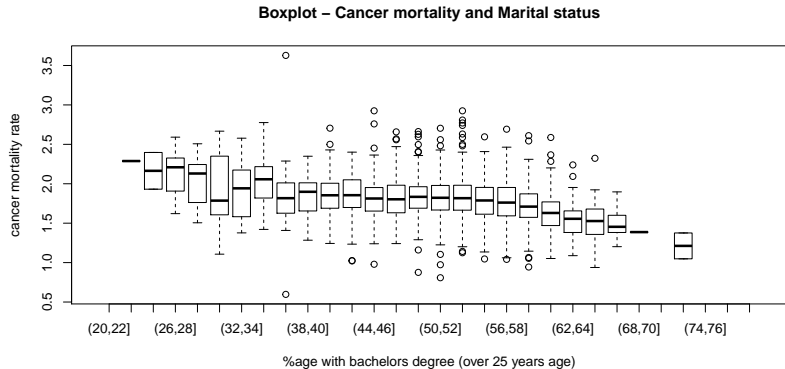Bottom 15 States by mortality
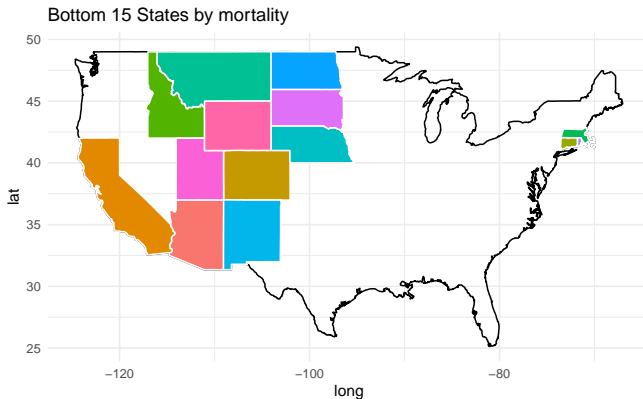
- ▶ Again, seems to be something in terms of the environment or the density of population.

Cancer Occurrences

Cancer prevelance and related mortality rate

Cancer occurence (per 1000)

Cancer related death rate (per 1000)

Marital Status & Household Size

Boxplot – Cancer mortality and Marital status

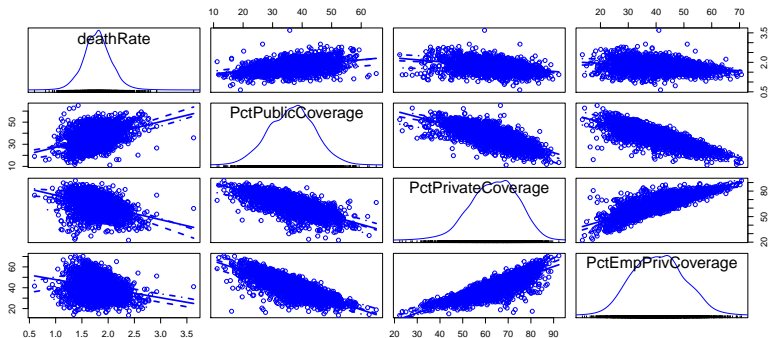# Post Secondary Education

Cancer Mortality Rate and Post Secondary Education
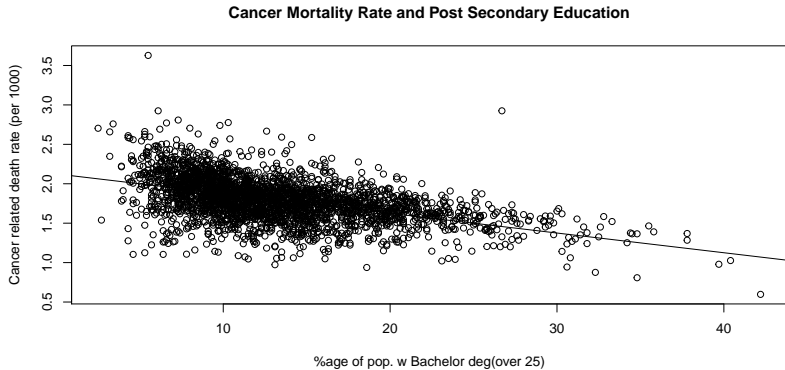
Insurance

Scatterplot Matrix for Insurance Types

Conclusion

Policy Advice

- Encourage people to seek higher education
- Encourage cohabitation
- Improve the public healthcare system
- Maintain low unemployment
- Address the issue of poverty