Crimes Dataset

Sociology 312, University of Oregon
Prof. Gullickson

Overview

The crimes data contain information on crime rates and demographic variables for all fifty US states and the District of Columbia. The crime rates are for the year 2010 and come from the FBI's Uniform Crime Reports (UCR). The UCR is a program where local law enforcement agencies all report crime statistics to the FBI and these are aggregated into final crime statistics. For our purposes, we are dividing crimes into two main categories of violent and property crime.

The demographic characteristics come from the American Community Survey (ACS) between the years 2008 and 2012. The ACS is an annual sample of the US population. To get a large enough sample in each state to calculate correct statistics (with little sampling error), I combine five years of data that are "centered"" on 2010.

Table 1 provides summary statistics for all the variables in the crime dataset.

Table 1: Summary statistics for all quantitative variables in the crimes dataset

| Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Violent	51	384.5	195.3	122.0	259.6	320.8	473.1	1330.2
Property	51	2894.0	641.5	1768.5	2406.1	2774.7	3441.4	4778.9
MedianAge	51	37.5	2.3	29.3	36.3	37.4	38.8	42.8
PctMale	51	49.3	0.8	47.3	48.8	49.2	49.7	52.0
PctLessHS	51	12.7	3.3	7.9	9.9	12.1	15.5	19.2
MedianIncomeHH	51	53.3	8.7	38.9	47.2	51.4	58.7	73.0
Unemployment	51	8.5	1.9	3.4	7.1	8.5	9.8	12.6
Poverty	51	14.3	3.1	8.4	12.1	14.2	17.0	22.3
Gini	51	45.5	2.2	41.3	44.0	45.6	46.8	53.1

Variable Descriptions

Violent

The violent crime rate. This includes the crimes of murder, rape, robbery, and aggravated assault. Both of our crime rates are determined by taking the number of crimes and dividing by the population of the state and then multiplying by 100,000 like so:

$$\frac{crimes}{population} * 100,000$$

The multiplication by 100,000 is just a convenience because otherwise the crime rates are very small numbers. We can interpret our crime rate as the number of violent crimes per year per 100,000 population.
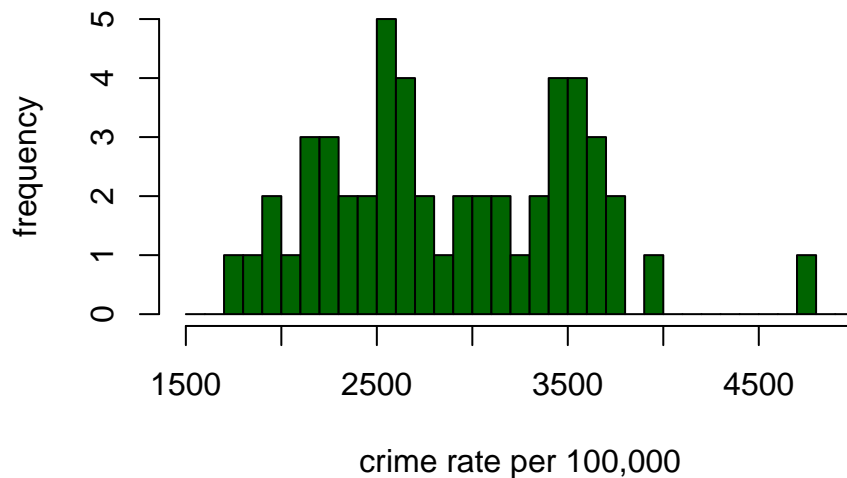
histogram of violent crime rate
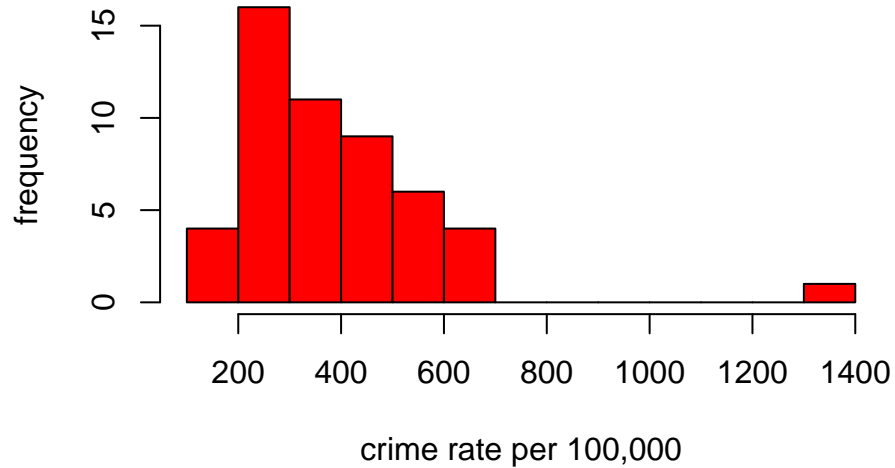
Property

The property crime rate. This includes the crimes of burglarly, larceny, and motor vehicle theft. Like the violent crime rate, the property crime rate can be intrepreted as the number of property crimes per year per 100,000 population.

histogram of property crime rate

MedianAge

The median age of the population in the state. This is potentially important because we know that criminal behavior is strongly related to age. People commit far less crime as they get older. Can you guess what the youngest state is? (hint: populations tend to be younger when fertility is higher)
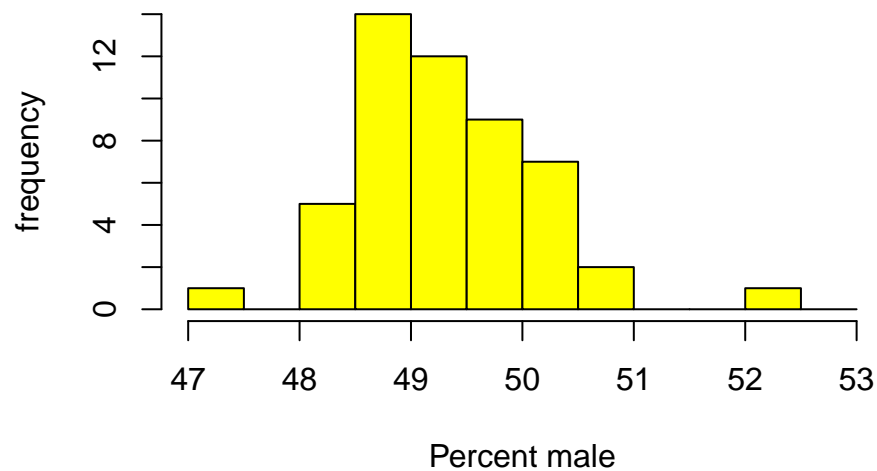
histogram of median age

frequency vs. median age
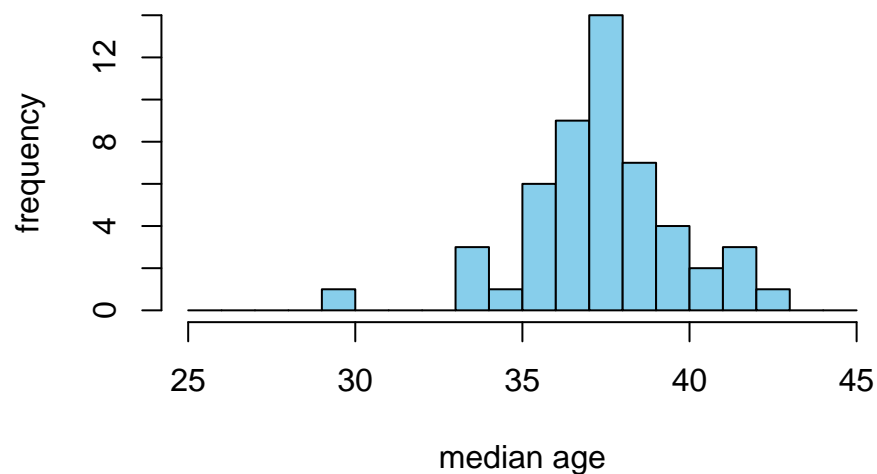
PctMale

The percent of the state population that is male. We know that men commit more crime than women, so this may be a useful variable.

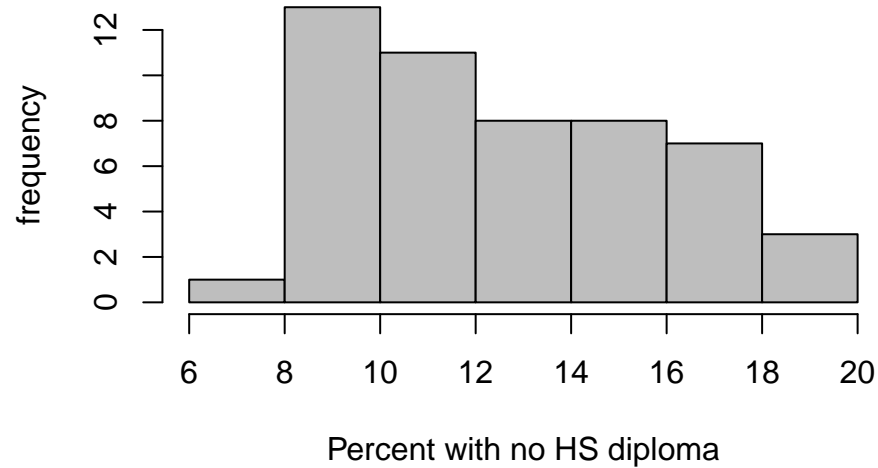## histogram of percent male

frequency vs. Percent male

PctLessHS

The percent of the state population over the age of 25 without a high school diploma. We limit to age 25 here to allow for people who may still be in the schooling process.

histogram of percent less than HS

Percent with no HS diploma

MedianIncomeHH

Median household income in the state. This is measured in thousands of dollars (i.e. 35 means $35,000). We are taking the income of each household (meaning all members of that household combined) rather than individual level income. For most purposes, this is thought to be a better measure because consumption and savings are typically organized at the household level.

histogram of median household income

Median household income
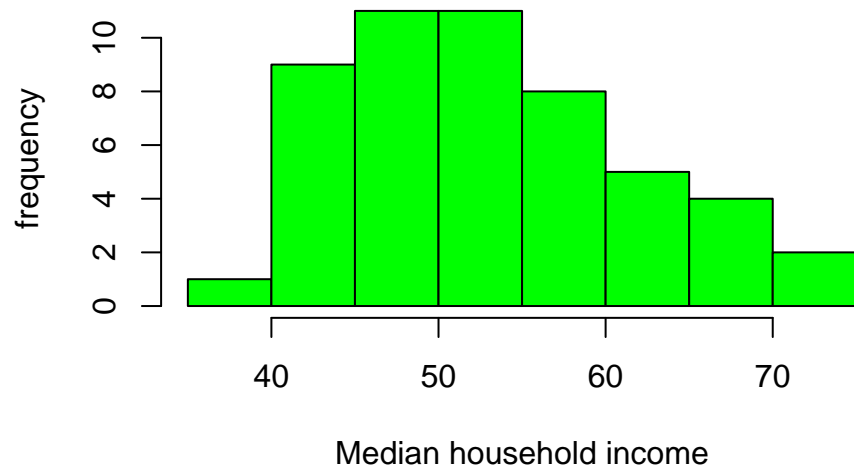
Unemployment

The unemployment rate in the state. The unemployment "rate" is really just a percentage. Its the percentage of individuals who are not working but want to work among all those in the labor force (those who are working or looking for work).

histogram of unemployment rate

frequency vs Unemployment rate (percent)
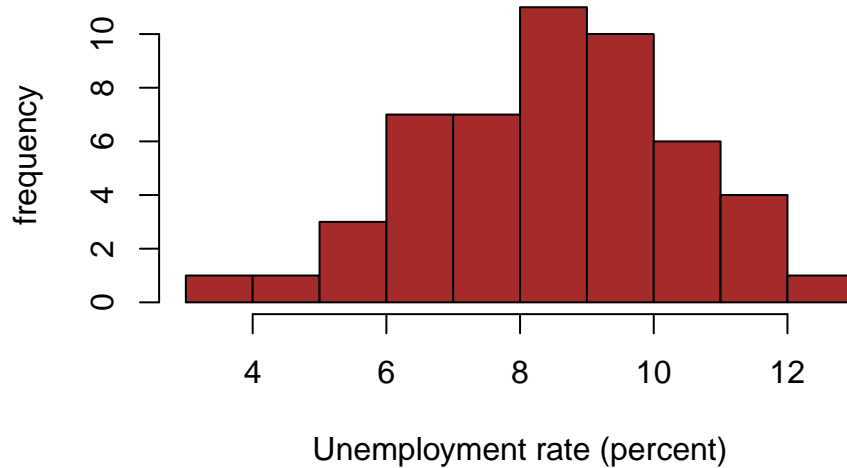
Poverty

The individual poverty rate in the state. The poverty "rate" is also really just a percentage. It is the percent of individuals living below the poverty line. The poverty line is a number developed by the federal government. It was originally developed in the 1960s and is adjusted for inflation every year. Many people critique the poverty line as being too low because it has not kept pace with increases in the consumer price index.

histogram of poverty rate

frequency vs Poverty rate (percent)
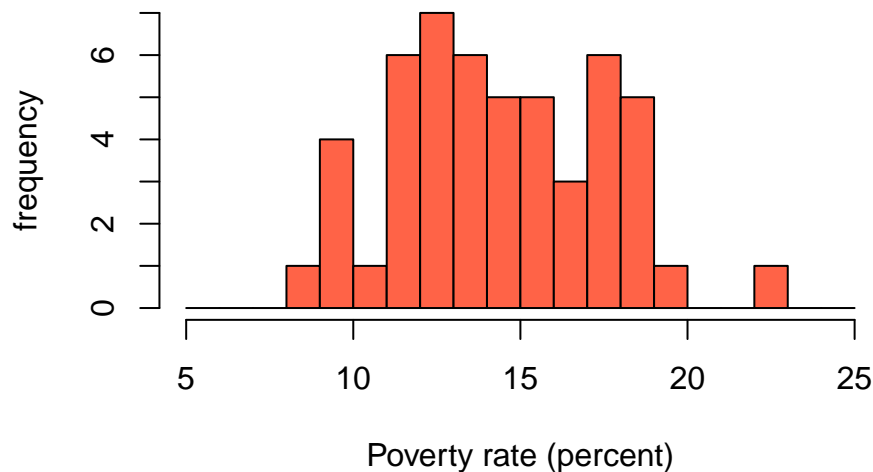
Gini

The gini coefficient which measures income inequality in a state. The gini coefficient is a widely used measure of how unequally income is distributed. If gini is zero, then everyone has exactly the same income. If gini is 100, then one person makes all the money and everyone else zero. The higher the gini coefficient, the more income inequality exists. You can get a more detailed description here.

histogram of the gini coefficient