

# LBW tutorial

*Joe Brew*

*10/01/2014*

For this activity, we'll be examining the incidence of low birth weight around the state of Florida. Data are from Florida Charts (<http://www.floridacharts.com>), a site well worth checking out for lots of aggregated health data in our state.

## 1. Get the data.

Create a folder named `lbw_tutorial` and download the data into that folder. The specific dataset for this activity is at [https://github.com/joebrew/misc/blob/master/lbw\\_tutorial/lbw.csv](https://github.com/joebrew/misc/blob/master/lbw_tutorial/lbw.csv). If you're familiar with github, you can clone the `joebrew/misc` directory entirely; if not, simply go to the previous link and download the data as a csv file (right click on the "Raw" button and save as a file).

## 2. Start an R script.

In the `lbw_tutorial` directory, create an R script named `lbw.R`.

## 3. Read in the data.

You'll do this in two steps. First use

```
setwd()
```

to set your working directory, then use

```
read.csv()
```

to read in the data. Assign it to an object named `lbw`.

## 4. Examine the data's structure.

Ensure that numerical / character / categorical variables are as they should be. Do this using

```
str()
head()
tail()
summary()
```

If not, transform them using

```
as.character()
as.numeric()
as.factor()
```

## 5. Explore the data's contents

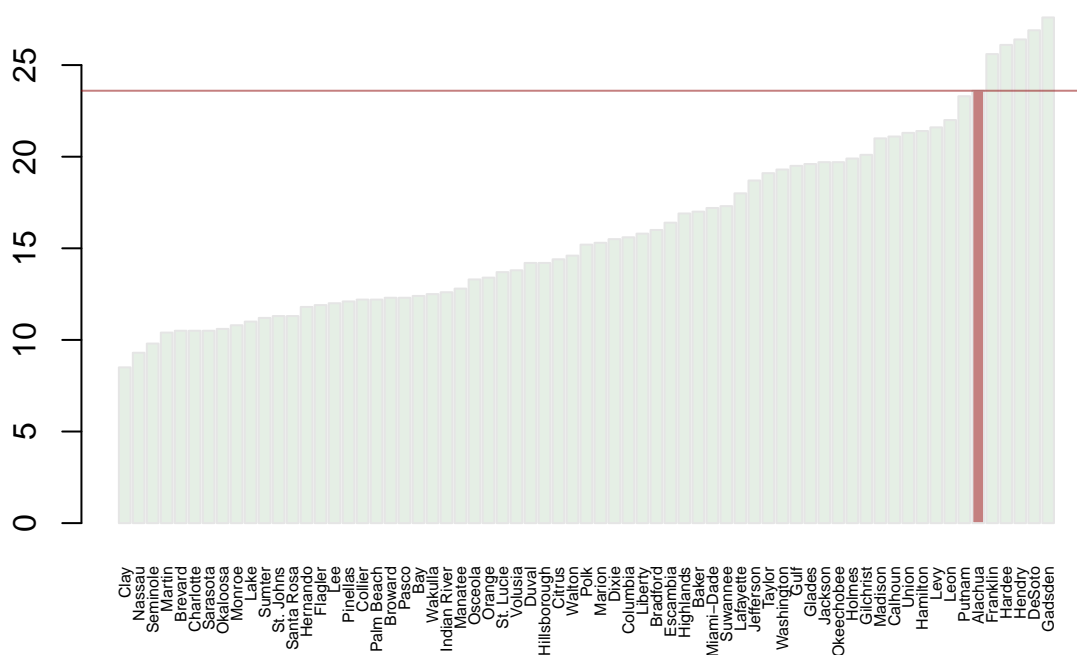
First, use dplyr to arrange (sort) the data by percent poverty

```
library(dplyr)
lbw <-
  lbw %>%
  arrange(percent_poverty)
```

Make an ordered barplot of the percent\_poverty variable using

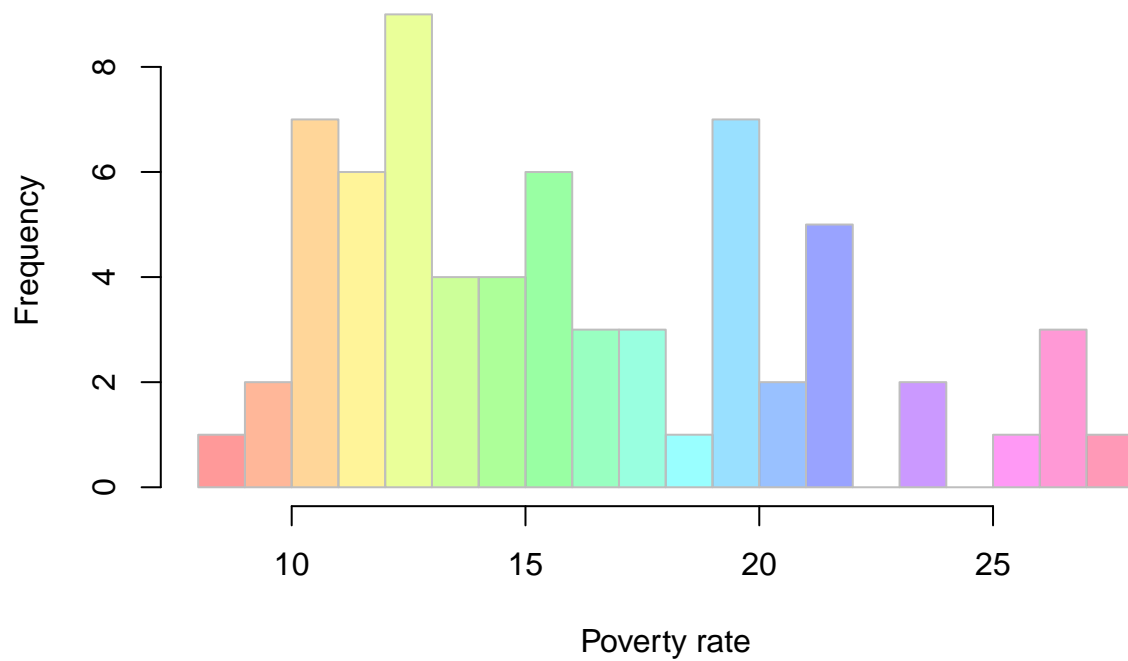
```
barplot()
```

Use the names.arg, cex.names and las = 3 arguments in the barplot() function to print the county names vertically below the chart. If you want a challenge, color Alachua's bar red and print a horizontal line at Alachua's poverty rate.



Make a histogram of percent\_poverty.

## Distribution of poverty rate in Florida by county



Re-arrange our data by population

```
lbw <- lbw %>%  
  arrange(total_pop)
```

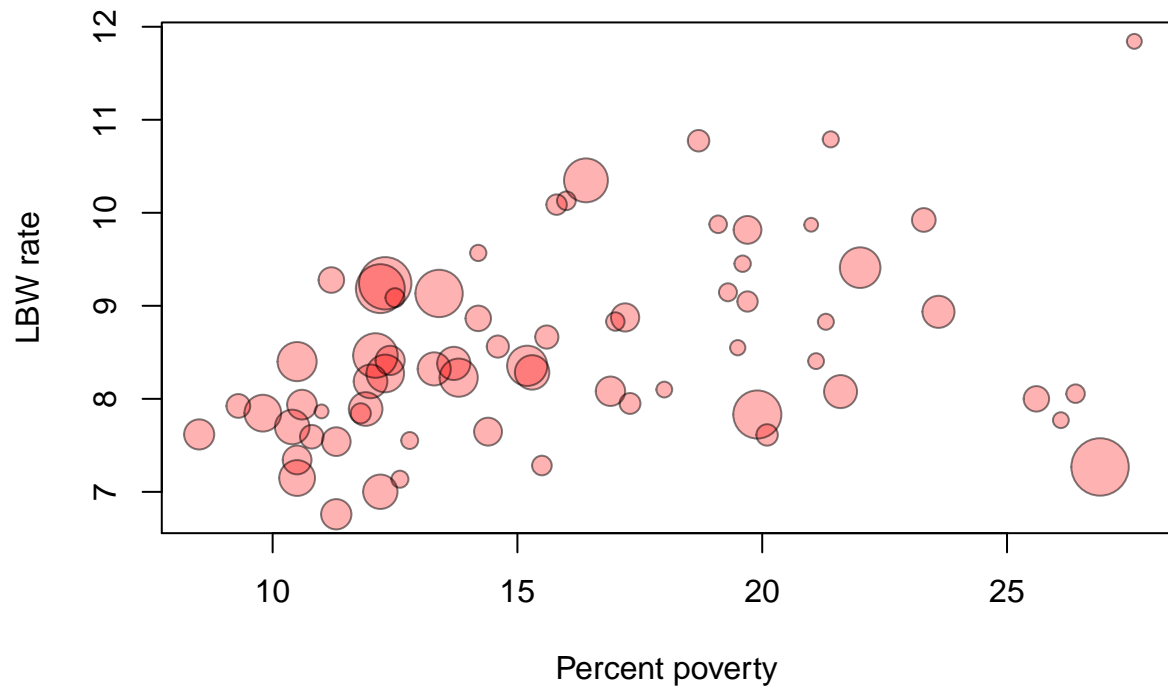
### 6. Create some useful new variables

Create variables named `white_per`, `black_per` and `other_per`. Also create `lbw_per`. These should be percent variables and should be easy to construct using simple division and the `total_pop` vector and `live_births` vector.

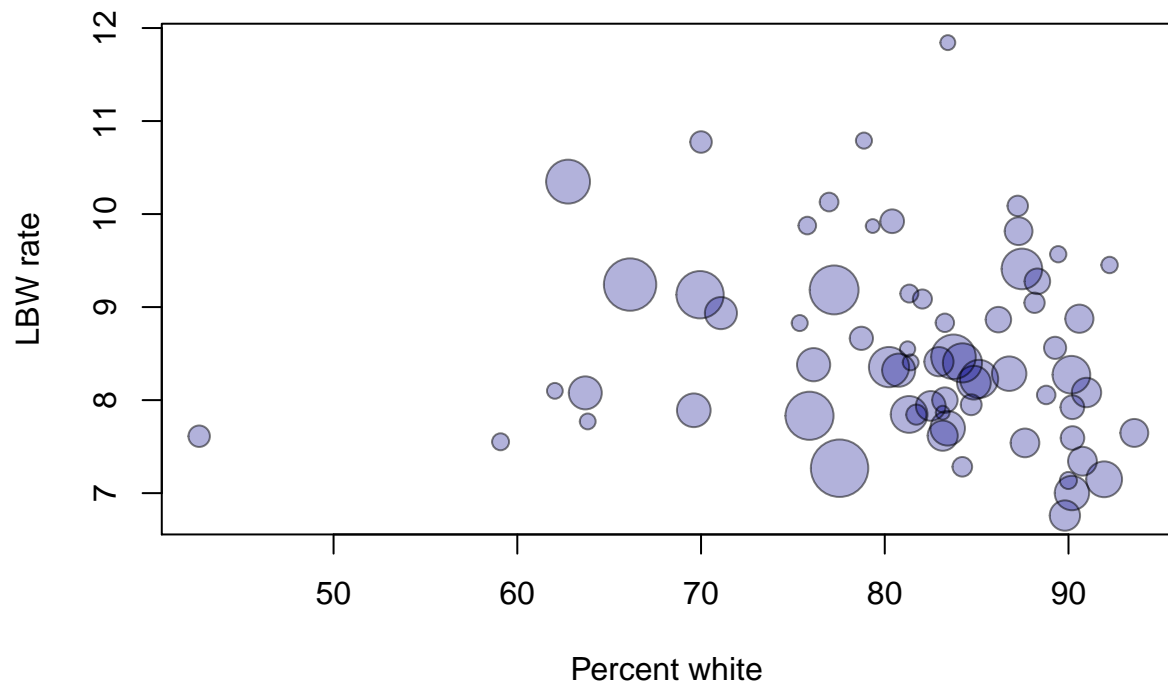
### 7. Plot the relationship between the poverty rate and lbw rate. Use

```
plot(x, y)
```

If you want to get fancy, make the size (`cex`) a function of a county's population size, as in the below plot



8. Plot the relationship between the percent of residents who are white and the lbw rate.



9. Construct a linear model to describe the relationship between lbw rate (y) and two independent variables: percent of residents who are black and the percent of residents who are poor. Name your linear model “fit”, and then summarize, plot it and calculate confidence intervals

```
fit <- lm()  
summary(fit)  
plot(fit)  
confint(fit)
```

10. Interpret the regression. At the county-level, after adjustment, what factors are significantly associated with lbw rate?