

Data Visualization with ggplot()

STAT 218 Lab Worksheet

Your Name

January 29th, 2024

Warning

We will learn how to visualize our data set today. You do not need to memorize these codes in this lecture. Please try to understand these conceptually.

Data Visualizations

Visuals are graphical representations of data. We use different colors, shapes, and the coordinate system to (1) summarize data; (2) tell a story or (3) explore data before conducting inferential statistics.

Let's start with loading today's data set by using `library()`, `data()`, and `glimpse()` function to start.

```
library(tidyverse)
library(openintro)
data(babies)
babies <- babies %>%
  mutate(smoke = as.logical(smoke),
         parity = as.logical(parity))
```

Before starting, let's have a look our data and variable types by using `glimpse()` function

```
# Put your glimpse function here
```

Let's use `?` to get more info

```
? babies
```

```
starting httpd help server ... done
```

You can see the info below in your Help tab.

case: id number

bwt: birth weight, in ounces

gestation: length of gestation, in days

parity: binary indicator for a first pregnancy (0 = first pregnancy)

age: mother's age in years

height: mother's height in inches

weight: mother's weight in pounds

smoke: binary indicator for whether the mother smokes

Visualizing a Single Categorical Variable

Let's choose one categorical variable and try to produce a bar plot

```
# Put your R Code here to produce a bar chart
```

Visualizing a Single Numeric Variable

Let's use **bwt** variable which is a numeric variable indicating birth weight in ounces. Choose your color from this [link](#)

```
# Put your R Code here to produce a histogram
```

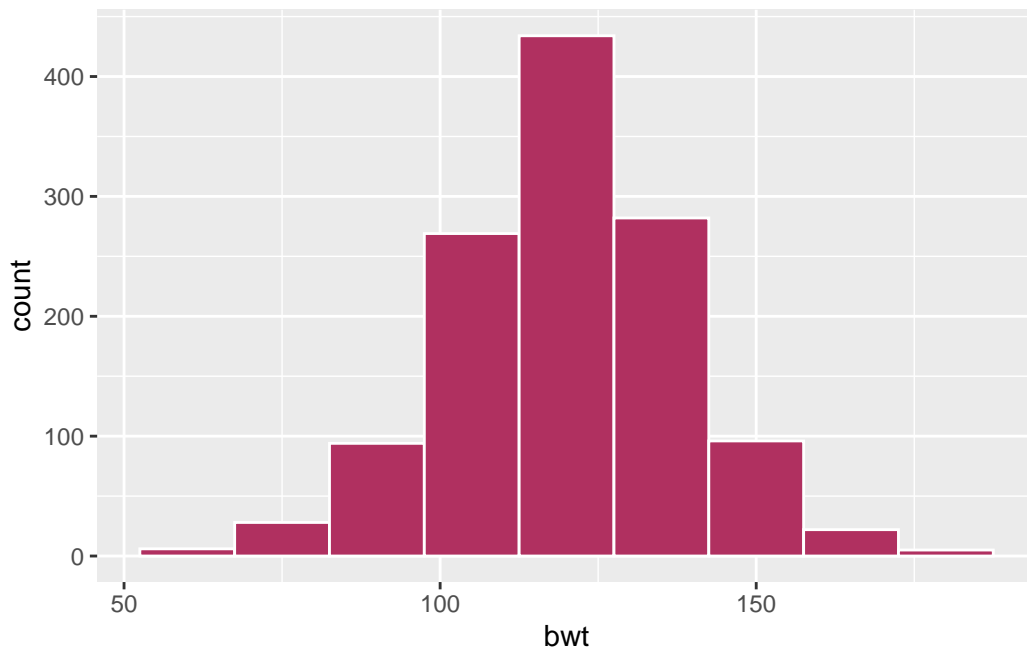
English

- Create a ggplot using the babies data frame.
- Map the **bwt** to the x-axis.
- Add a layer of a histogram.
- Change the binwidth to 15.
- Color the borders of the bars (bins?) as white.
- Fill it with a color code named maroon

R Code

```
ggplot(data = babies,  
       aes(x = bwt)) +  
  geom_histogram(binwidth = 15,  
                 color = "white",  
                 fill = "maroon")
```

Plot



Visualizing Two Categorical Variables

Now we will try to fill the y-axis as if it is something look like percentage which is called **Standardized Bar Plot**. Note that y-axis is no longer count but we will learn how to change that later.

```
# Put your R Code here to produce a standardized bar plot
```

Visualizing One Categorical and One Numeric Variable

We are visualizing a single numerical (`bwt`) and single categorical variable (`smoke`) by using `geom_boxplot`

```
# Put your R Code here to produce a box plot
```

Visualizing Two Numeric Variables

We are visualizing two numerical (`bwt` & `gestation`) by using `geom_point`

```
# Put your R Code here to produce a scatterplot
```

Visualizing More than Two Variables

We are visualizing two numerical (`bwt` & `gestation`) and one categorical (`smoke`) variables

```
# Put your R Code here
```