

Making plots with ggplot2: histograms, boxplots, line graphs

Emily Malcolm-White

```
# load packages
library(tidyverse)
```

possum data

The possum data frame consists of nine morphometric measurements on each of 104 mountain brushtail possums, trapped at seven Australian sites from Southern Victoria to central Queensland.

There are two different populations (pop): **Vic** (Victoria) and **other** (New South Wales or Queensland)

```
#Import the possum data again
library(DAAG)
data("possum")
```

Histograms

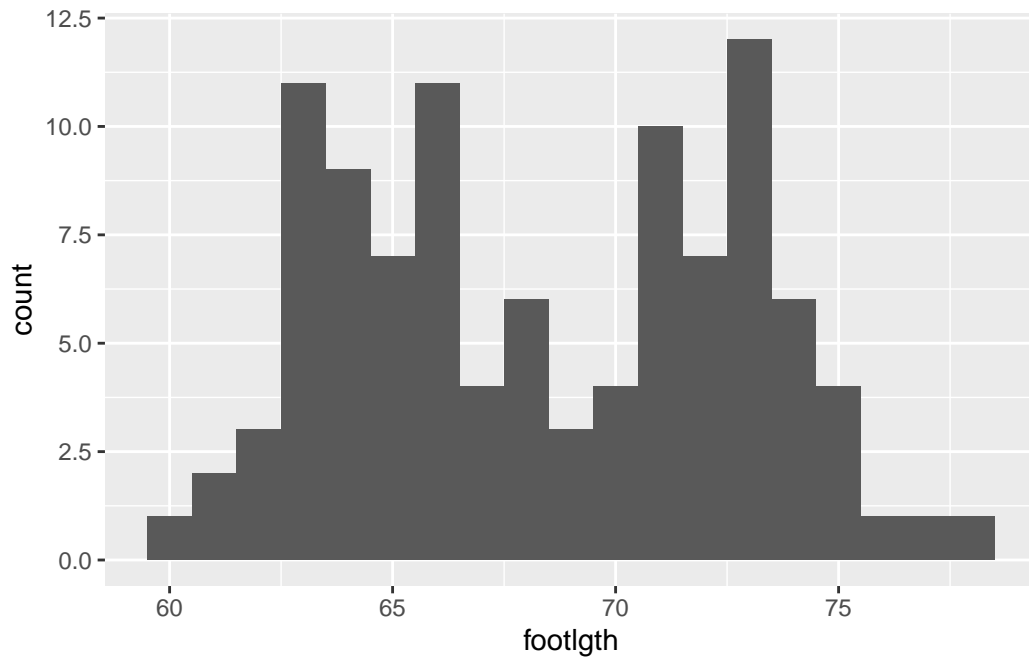
💡 Tip

Histograms are great for looking at the distributions of numeric variables

A boxplot for the footlength of all possums in this dataset:

```
ggplot(possum, aes(x=footlgth)) + #<1>
  geom_histogram(binwidth=1) #<2>
```

- ① only one x variable is needed
- ② you can adjust the binwidth, as needed



Boxplots

💡 Tip

Boxplots are good for displaying the spread, central tendency, and distribution of one numeric variable.

A lone box-plot for one numeric variable (foot length) with some custom colors:

```
ggplot(possum, aes(y=footlgth)) +  
  geom_boxplot(color="blue", fill="lightblue") #<1>
```

- ① The **geom** for boxplot. The **color** arguments makes the outline of the boxplot blue and the the **fill** argument shades the inside of the inner quartile range.

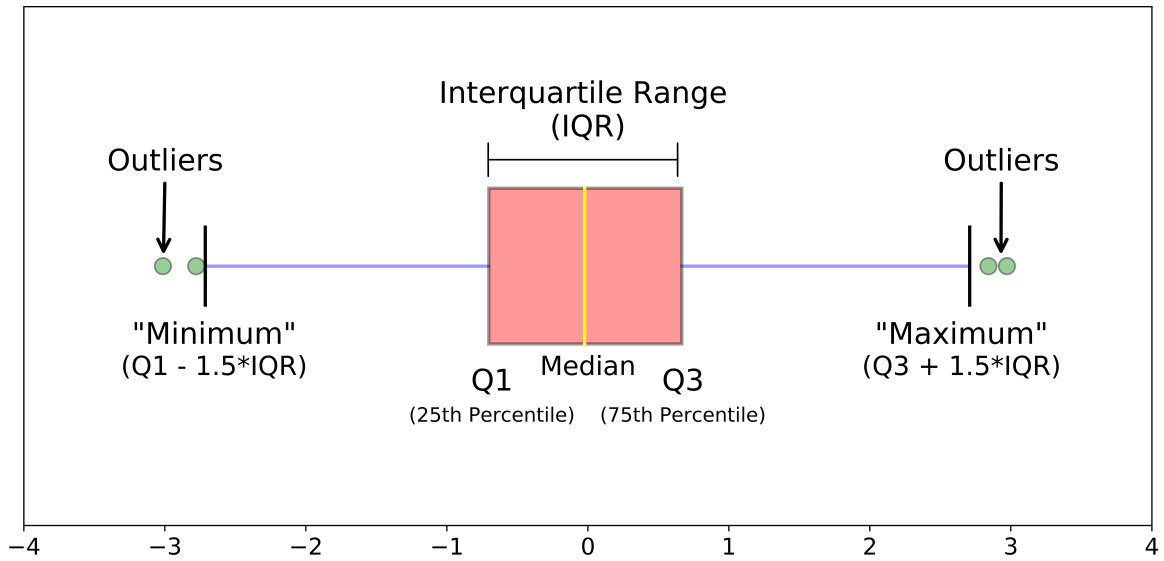
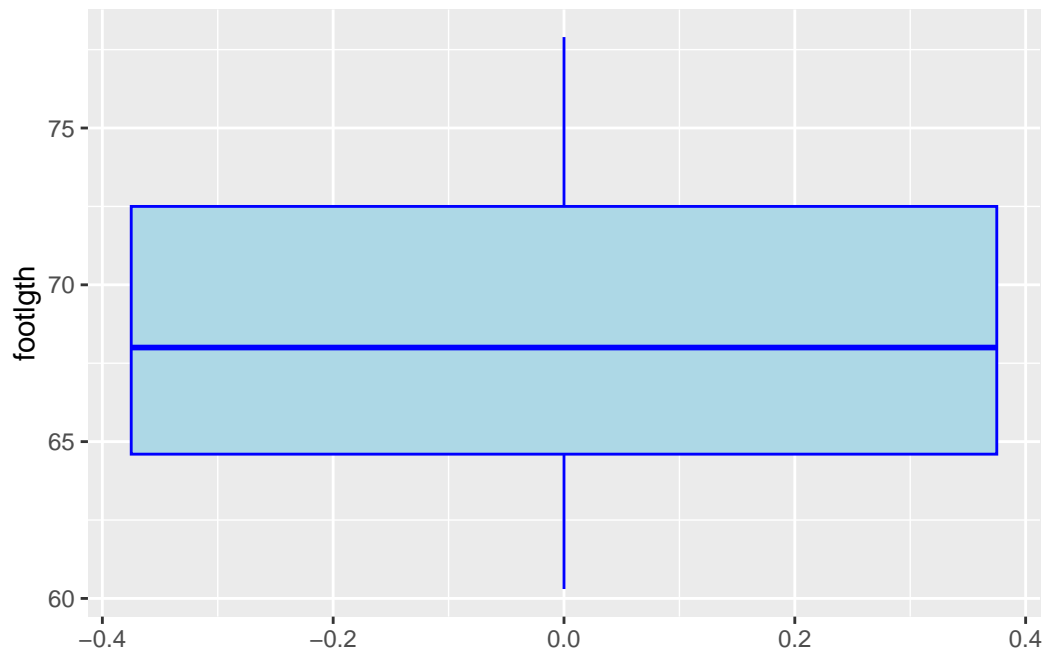
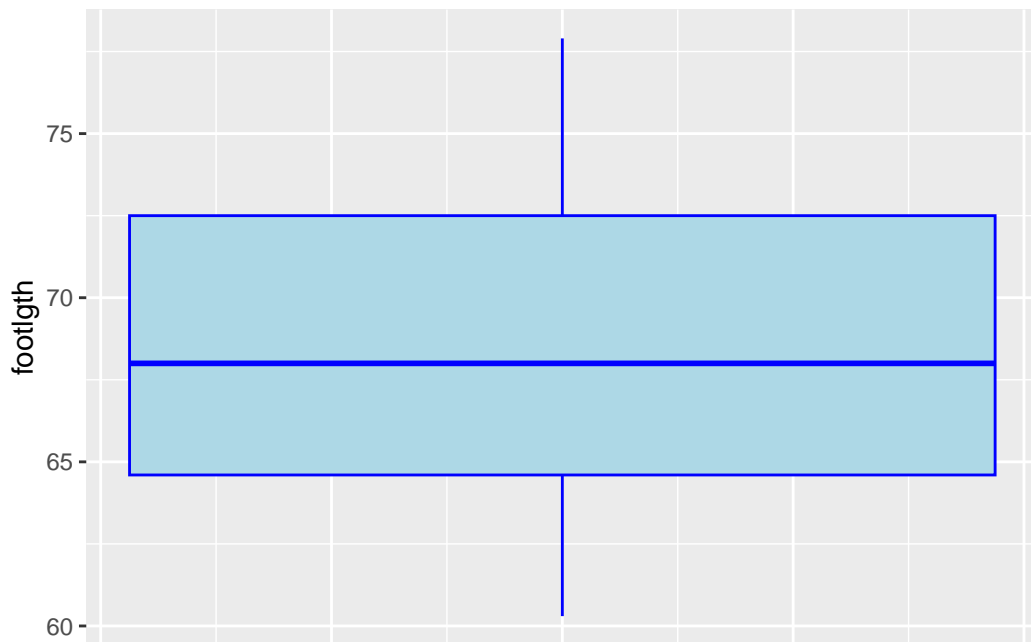


Figure 1: Credit: Michael Galarnyk



If only one boxplot, it puts weird numbers on the x axis, you may want to use the theme to hide these numbers.

```
ggplot(possum, aes(y = footlgth)) +
  geom_boxplot(color = "blue", fill = "lightblue") +
  theme(
    axis.text.x = element_blank(), # Hide text
    axis.ticks.x = element_blank(), # Hide tick marks
    axis.title.x = element_blank() # Hide axis title
  )
```

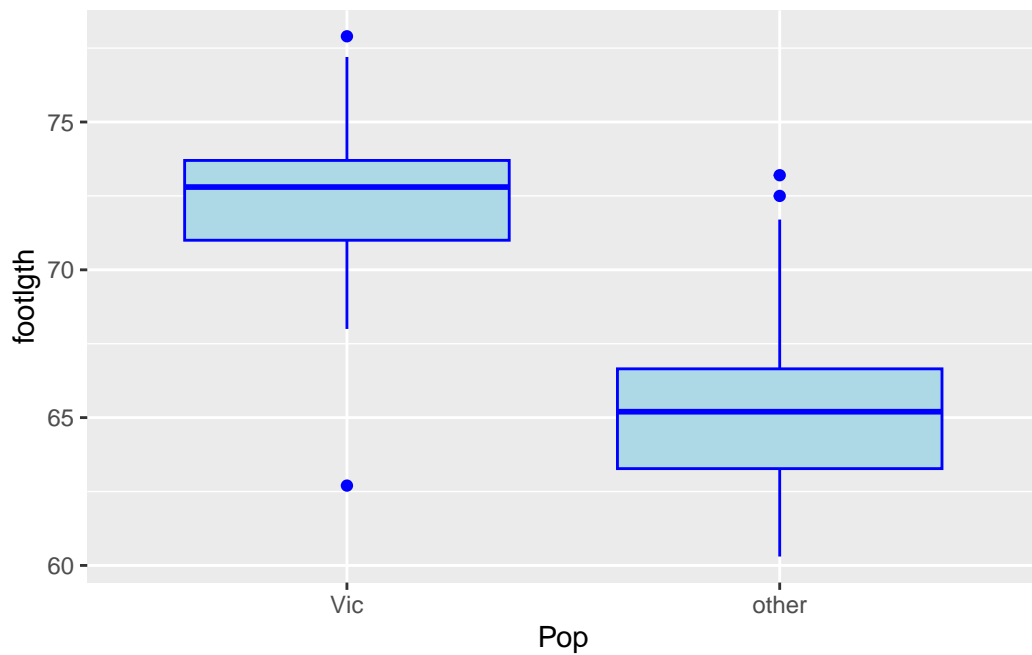


💡 Tip

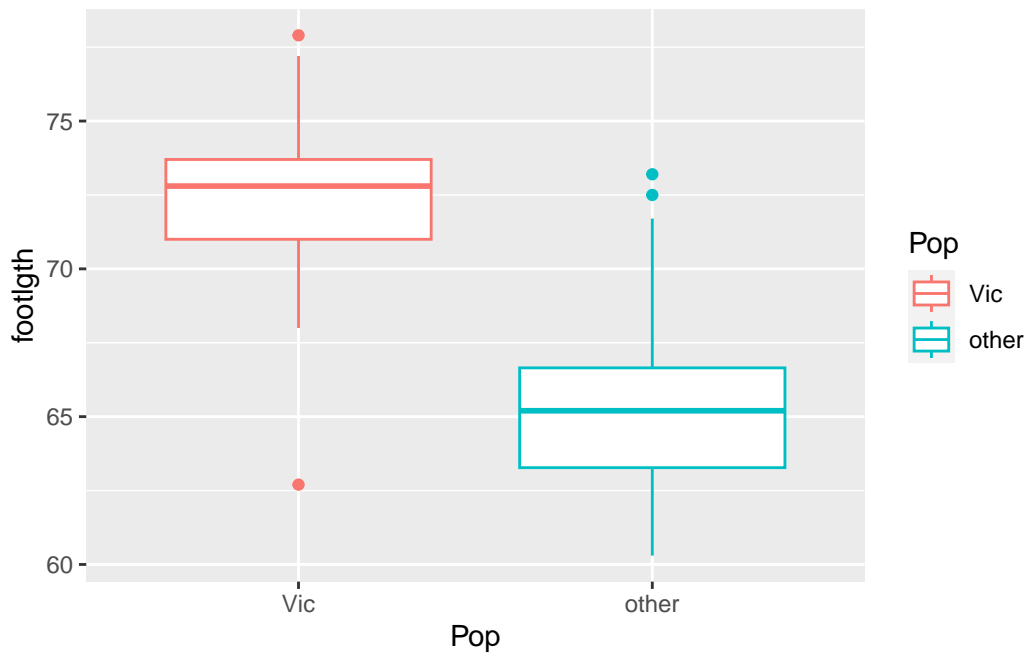
Side-by-side boxplots are good for displaying one categorical variable and one numeric variable. One advantage of boxplots over bar plots is that they are able to show a bit about the spread and distribution of the numeric variable!

A side-by-side boxplot to compare the foot lengths between the two populations of possums:

```
# same color for both boxplots
ggplot(possum, aes(y=footlgth, x=Pop)) +
  geom_boxplot(color="blue", fill="lightblue")
```



```
# different color for both boxplots
ggplot(possum, aes(y=footlgth, x=Pop, color=Pop)) +
  geom_boxplot()
```



Line Graph

This dataset was produced from US economic time series data available from <https://fred.stlouisfed.org/>. Type ? economics to learn more.

```
data("economics")
```

Create a line plot with the unemployment rate of the US over time:

```
# Create a line plot of unemployment over time
ggplot(economics, aes(x = date, y = unemploy)) +
  geom_line(color = "darkblue", size = 1) +
  labs(
    title = "U.S. Unemployment Over Time",
    x = "Year",
    y = "Number of Unemployed (in thousands)"
  )
```

