

A2Q3

Ian Murphy

2019-09-19

Question 3

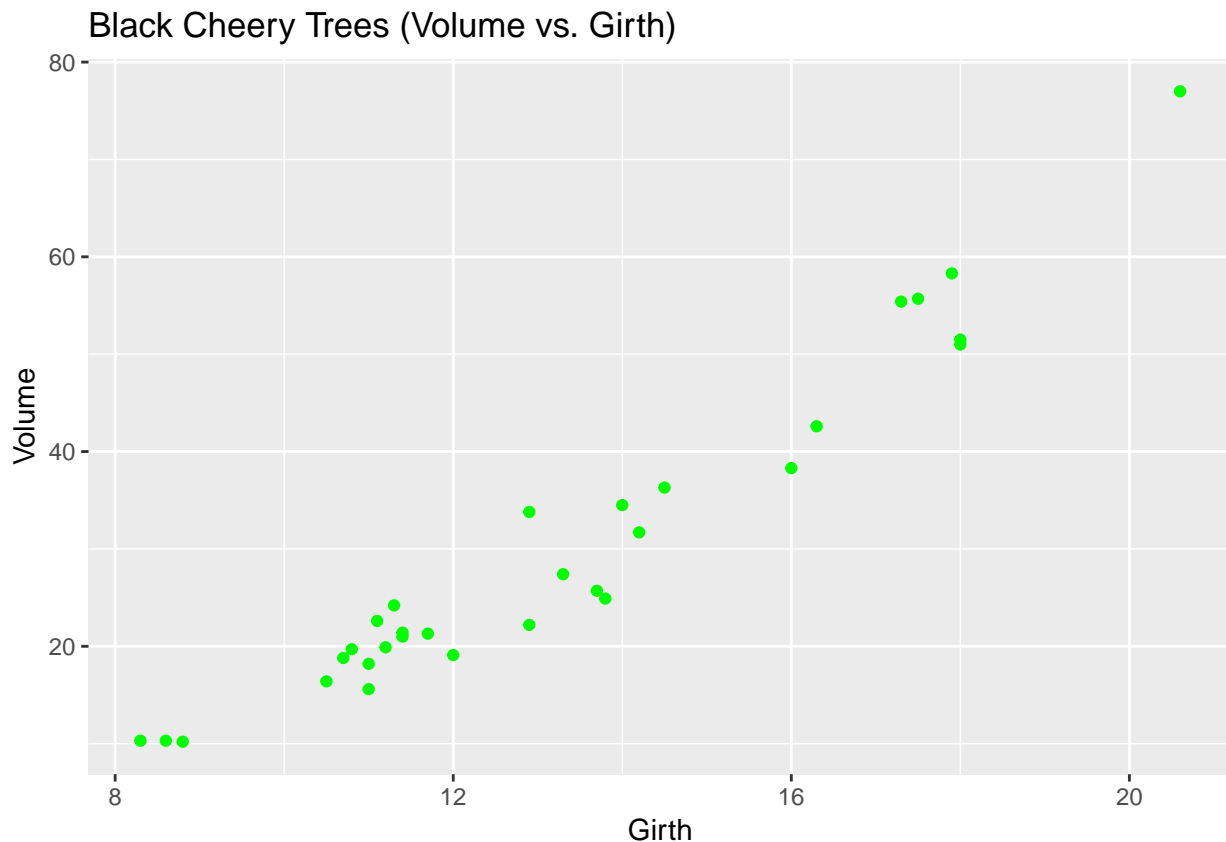
We will explore different plots. To do so, we will analyze the “trees” dataset from R. This dataset contains information from 31 black cherry trees, including their Girth, Height, and Volume.

We want to see if any of these variables are related. To do so, we will simply look at the scatterplots.

The first one will be a scatterplot of Girth vs. Volume, which clearly shows a nice linear trend.

```
library(tidyverse)
```

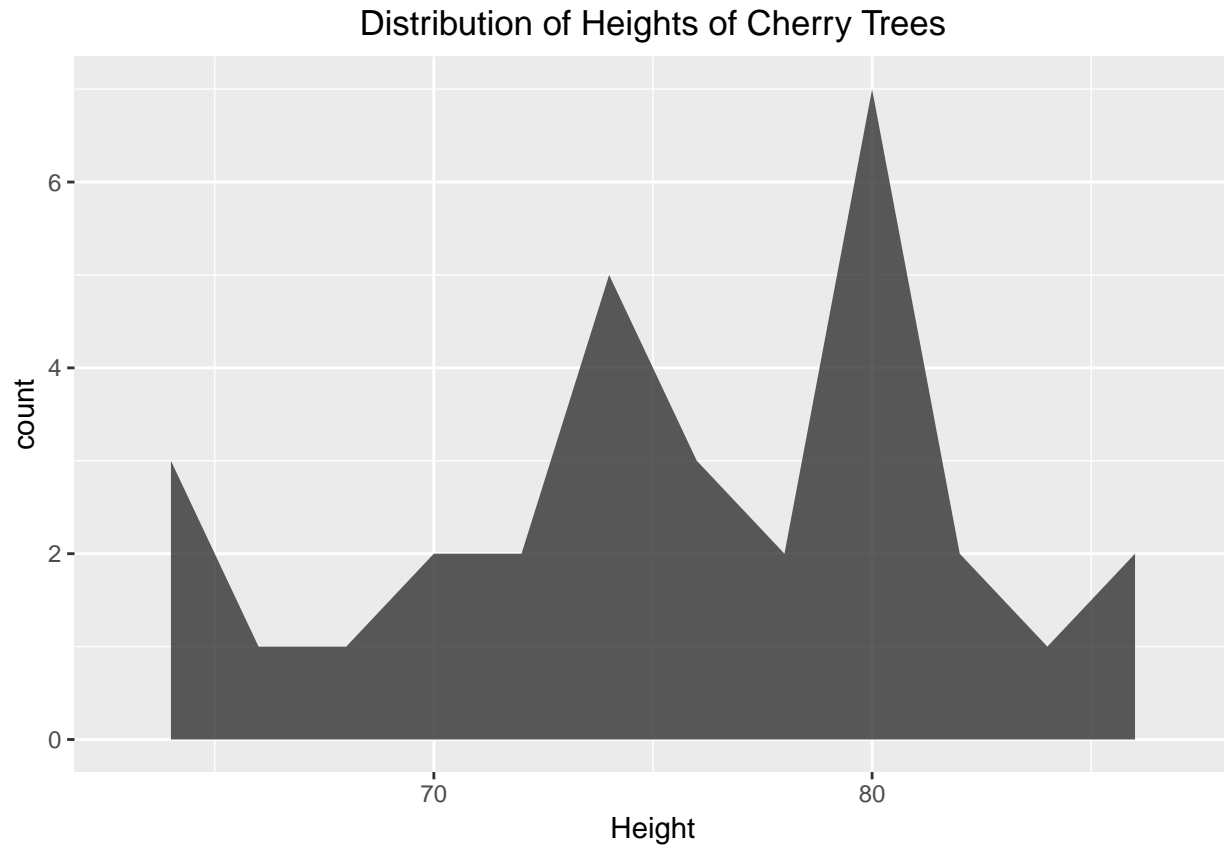
```
trees %>%  
  ggplot(aes(Girth, Volume)) +  
  geom_point(colour="green") +  
  ggtitle("Black Cheery Trees (Volume vs. Girth)")
```



There seems to be an upward trend for Volume and Girth.

For part 2, we wish to look at a different type of plot, rather than just scatterplots. Let's look at an area plot (similar to histogram) of the distribution of heights of cherry trees.

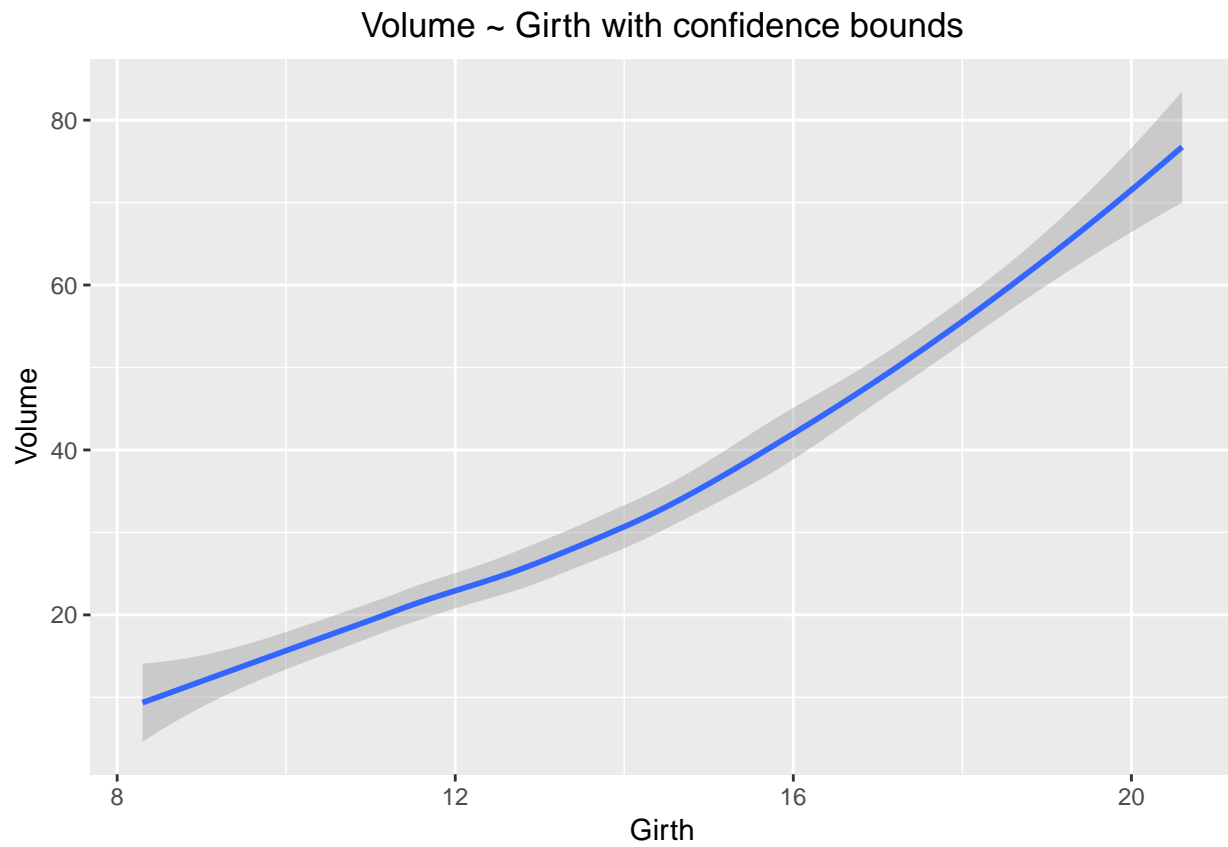
```
trees %>%
  ggplot(aes(Height)) +
  geom_area(stat="bin", alpha=0.8, binwidth=2) +
  ggtitle("Distribution of Heights of Cherry Trees") +
  theme(plot.title = element_text(hjust = 0.5))
```



How about one more plot, just for fun?!

```
trees %>%
  ggplot(aes(Girth, Volume)) +
  geom_smooth() +
  ggtitle("Volume ~ Girth with confidence bounds") +
  theme(plot.title = element_text(hjust = 0.5))
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Above, we predicted Volume using Girth. Then produced a LOESS curve, with appropriate confidence bounds on the plot.