

Data Visualization With Stata (The Basics)

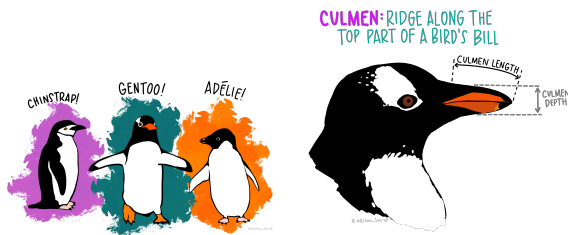
Andy Grogan-Kaylor

19 Dec 2020

Introduction

99% of data visualization work sometimes seems to consist of creating bar graphs (`graph bar y, over(x)`) and scatterplots (`twoway scatter y x`). (For the sake of completeness, I am also going to mention histograms (`histogram x`).) I am not a particular fan of Stata's default graph schemes, so I am going to make use of the graph scheme entitled `sicolor`.

This is a quick guide to these ideas using the Palmer Penguins Data.



```
. clear all

. use "https://github.com/agrogan1/newstuff/raw/master/data-visualization-with-Stata-the-basics/pe
> nguins.dta", clear

. set scheme sicolor // use sicolor scheme
```

Histogram: histogram

```
. histogram body_mass_g, title("Body Mass of Penguins")
(bin=18, start=2700, width=200)

. graph export myhistogram.png, width(1000) replace
(file myhistogram.png written in PNG format)
```

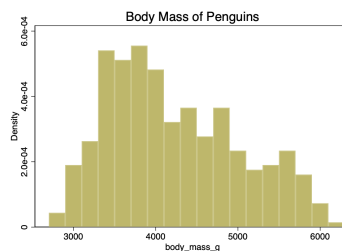


Figure 1: histogram

Bar Graph: graph bar

Counting Up Numbers In Each Group

```
. graph bar, over(species) title("Penguin Species")  
  
. graph export mybar1.png, width(1000) replace  
(file mybar1.png written in PNG format)
```

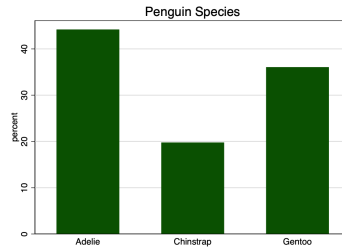


Figure 2: bar graph

Average Of A Continuous Variable Across Groups

```
. graph bar body_mass_g, over(species) title("Body Mass of Penguin Species")  
  
. graph export mybar2.png, width(1000) replace  
(file mybar2.png written in PNG format)
```

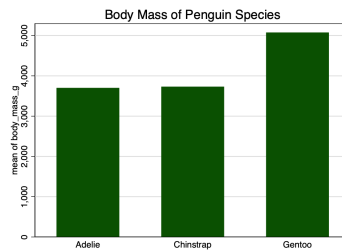


Figure 3: bar graph

Scatterplot: twoway scatter

```
. twoway scatter culmen_length_mm body_mass_g, title("Penguin Culmen Length by Body Mass")  
  
. graph export myscatter.png, width(1000) replace  
(file myscatter.png written in PNG format)
```

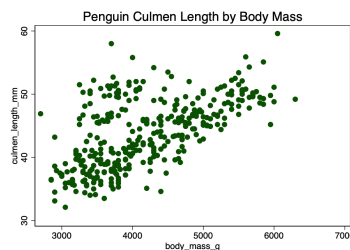


Figure 4: scatterplot