Data Visualization With Stata (The Basics)

Andy Grogan-Kaylor

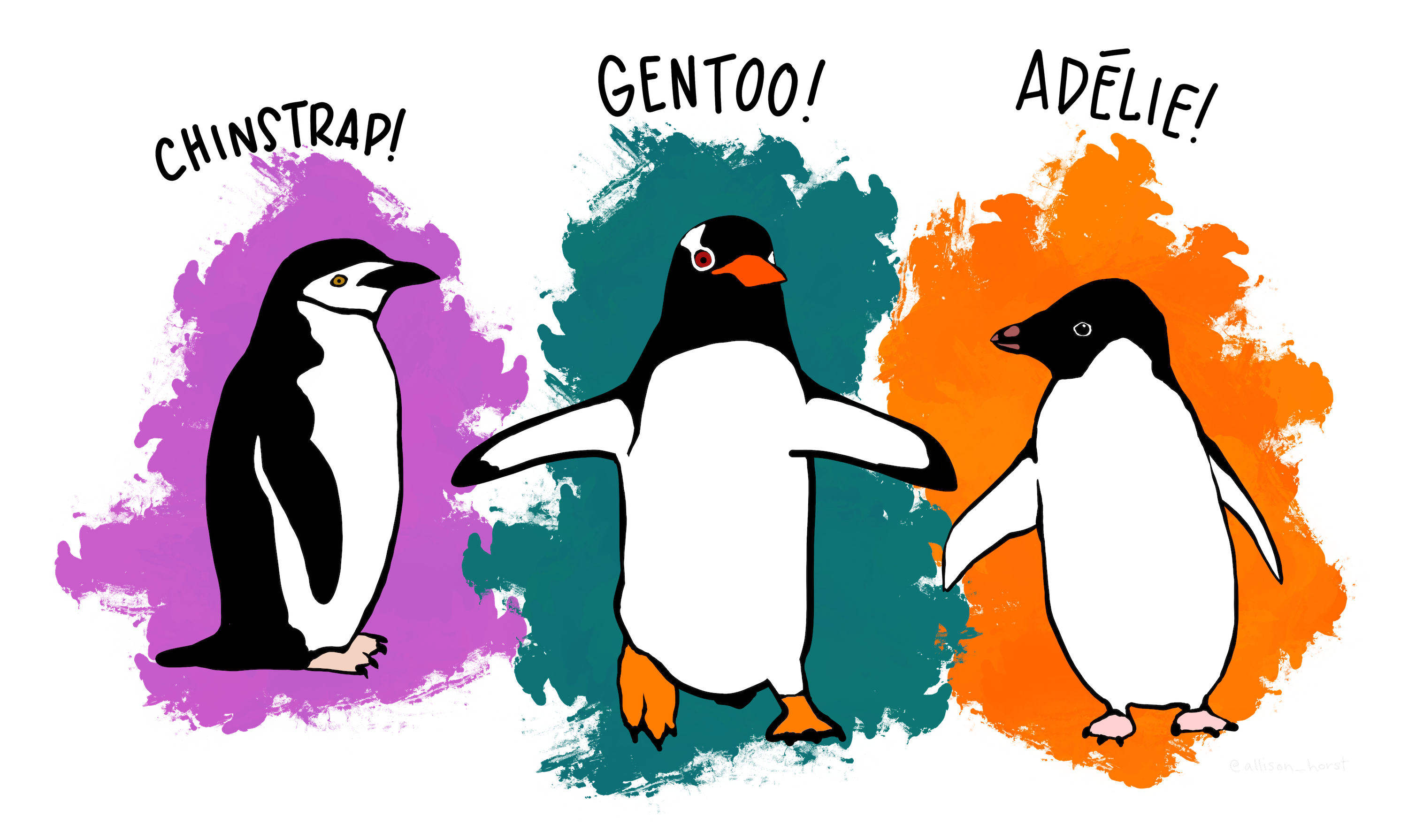
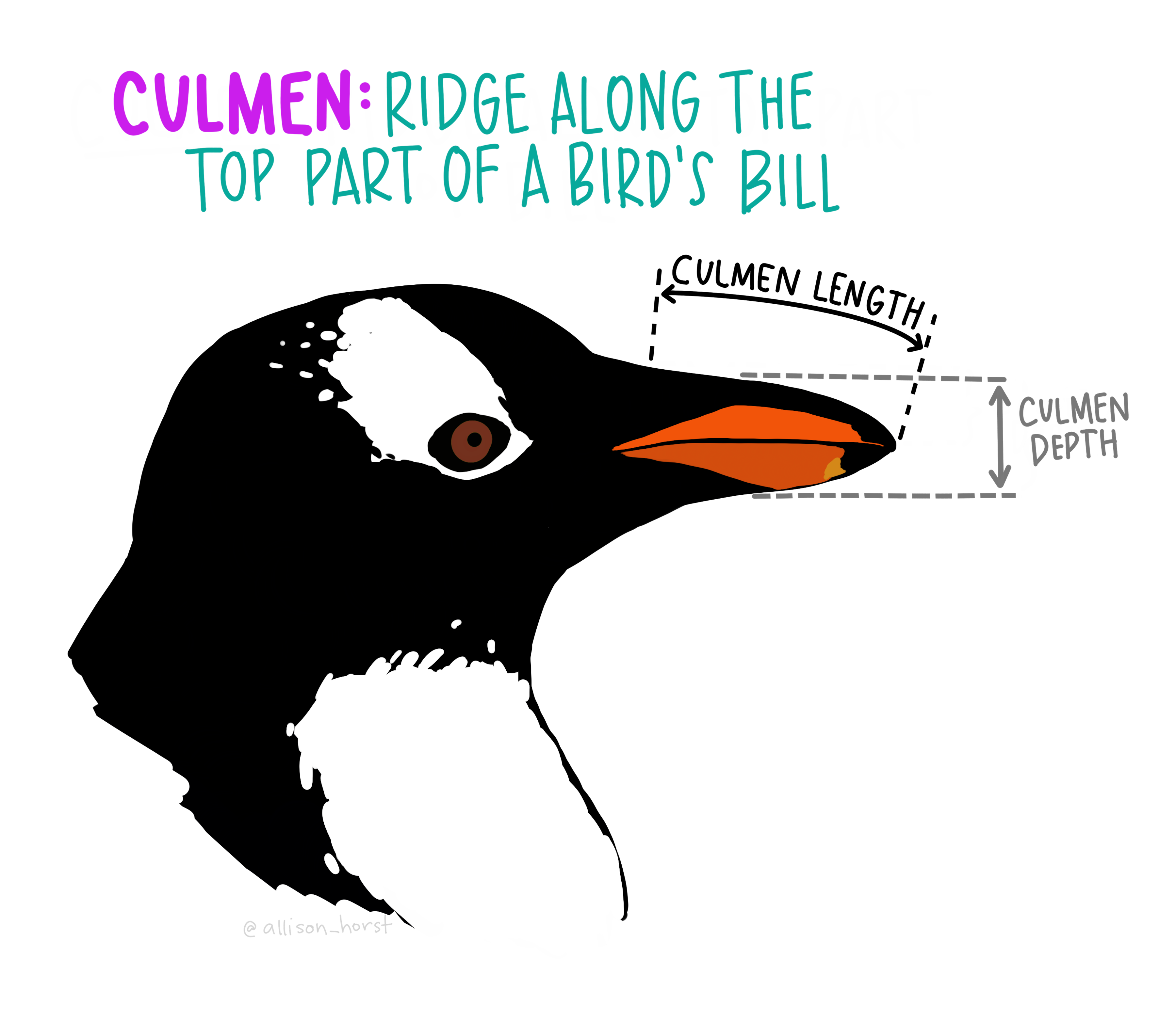
4 Feb 2022

# Introduction

99% of data visualization work 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).)

Note: In some commands, I use /// so that Stata commands can be on multiple lines.

This is a quick guide to these ideas using the [Palmer Penguins Data](https://github.com/allisonhorst/palmerpenguins/blob/master/README.md).

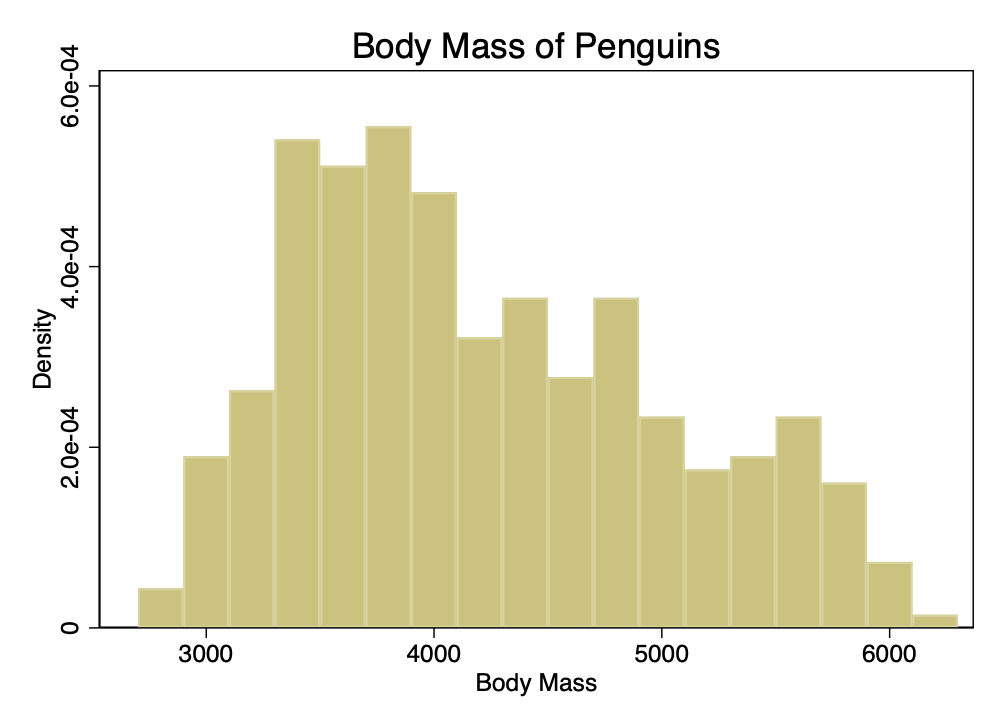
. clear all  
  
.   
. use "https://github.com/agrogan1/Stata/raw/master/data-visualization-with-Stata-the-basics/penguins.  
> dta", clear

I am not a particular fan of Stata’s default graph schemes, so I am going to make use of the graph scheme entitled s1color.

. set scheme s1color // use s1color scheme

# Histogram: histogram x

. histogram body\_mass\_g, title("Body Mass of Penguins") xtitle("Body Mass")  
(bin=18, start=2700, width=200)

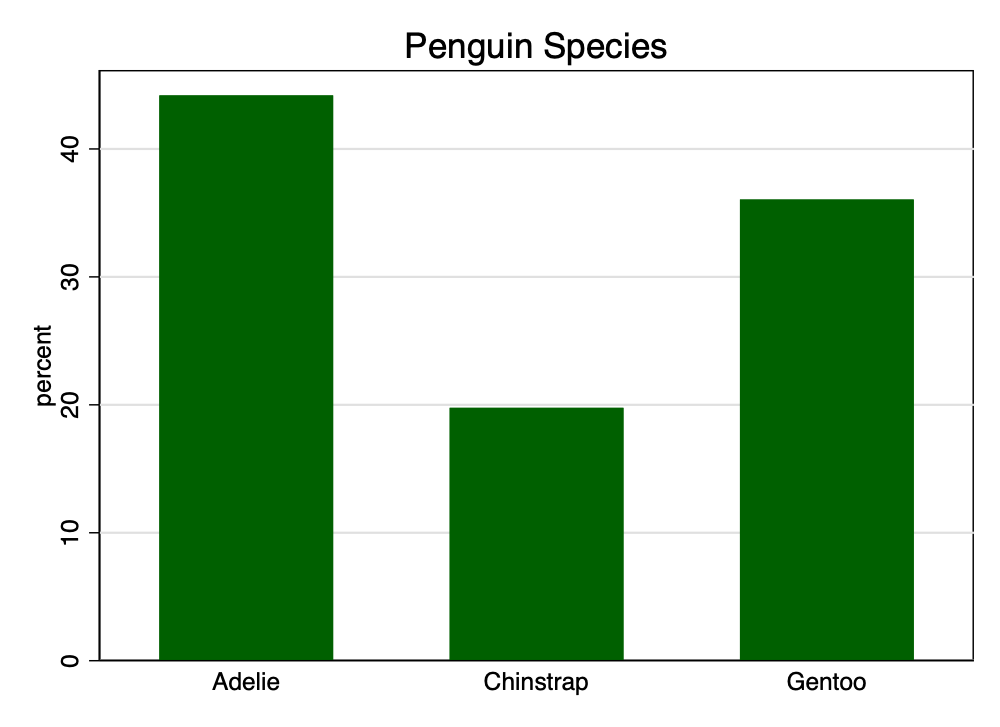


histogram

# Bar Graph: graph bar

## Counting Up Numbers In Each Group: graph bar, over(x)

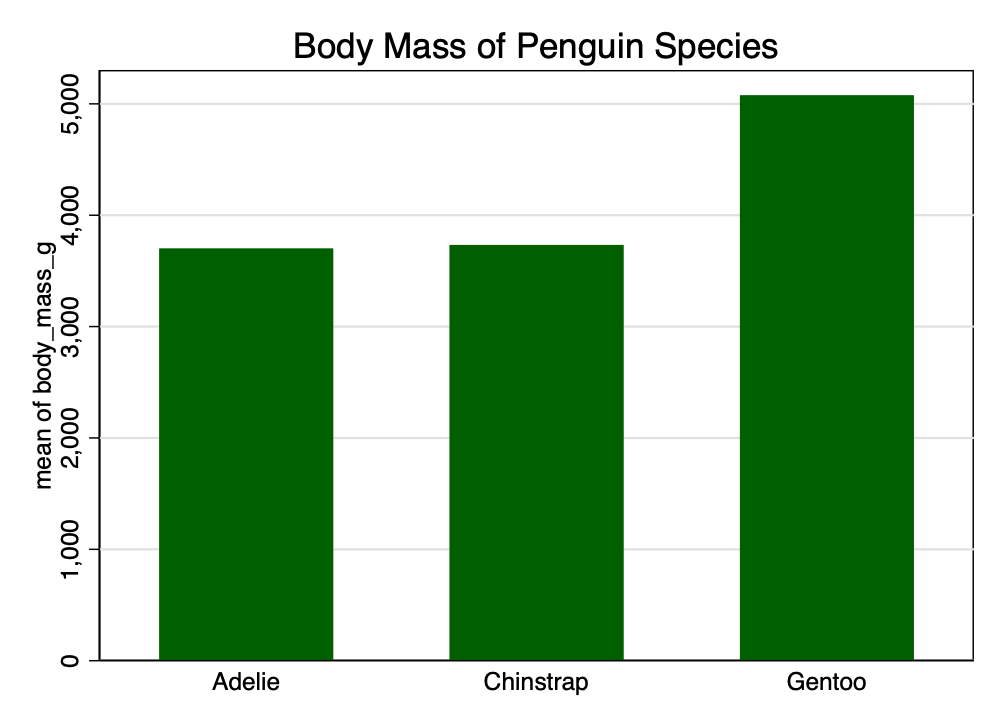
. graph bar, over(species) title("Penguin Species")



bar graph

## Average Of A Continuous Variable Across Groups: graph bar y, over(x)

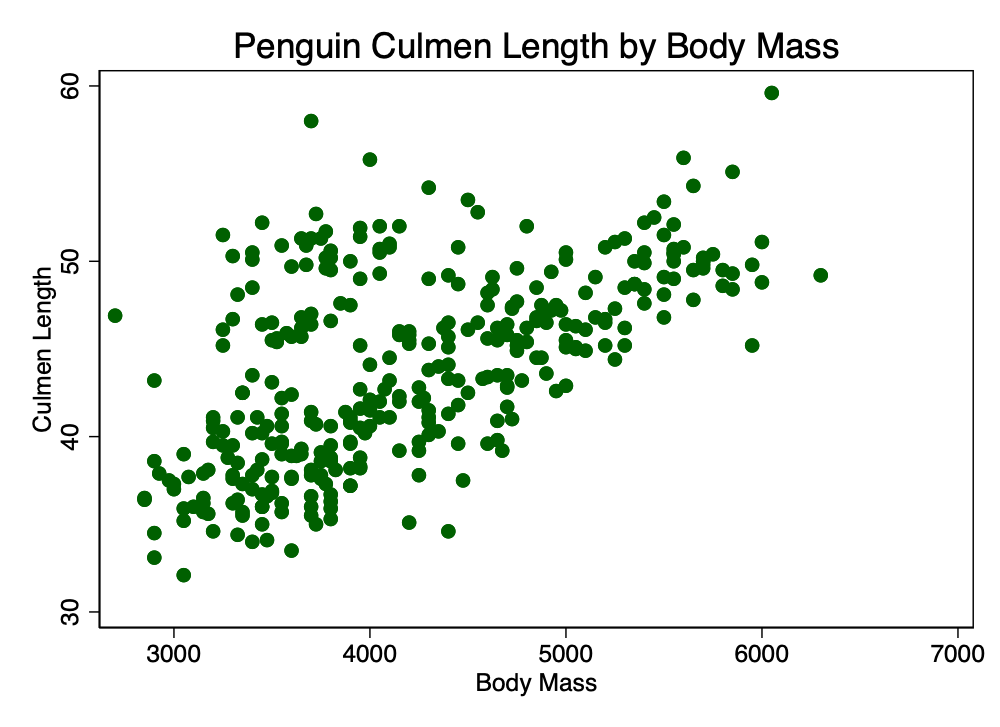
. graph bar body\_mass\_g, over(species) title("Body Mass of Penguin Species")



bar graph

# Scatterplot: twoway scatter y x

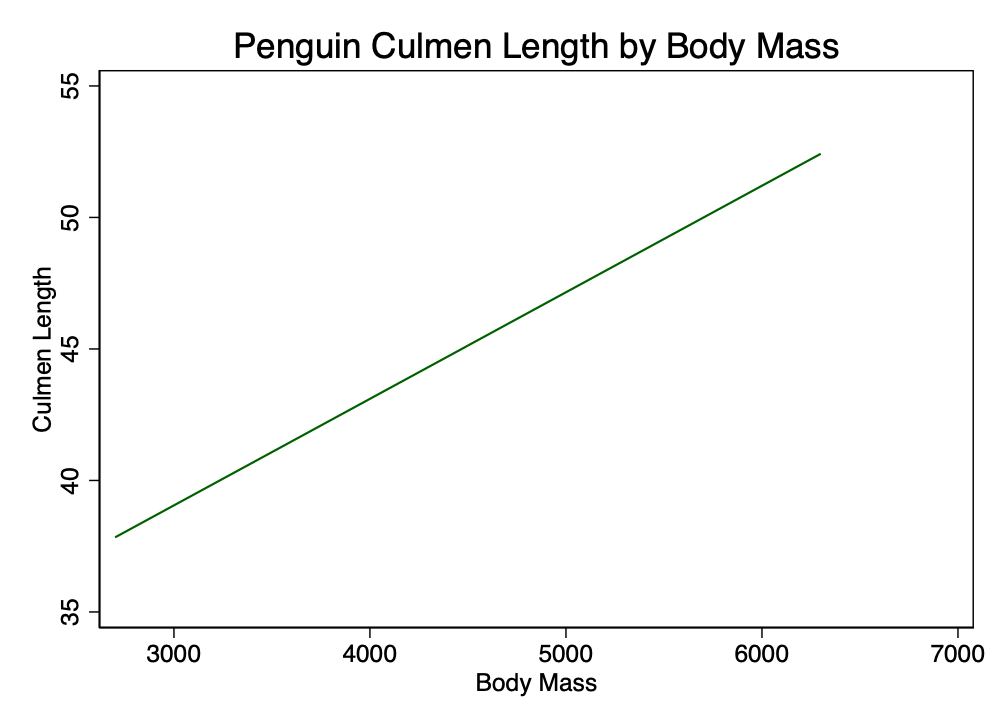
. twoway scatter culmen\_length\_mm body\_mass\_g, ///  
> title("Penguin Culmen Length by Body Mass") ///   
> xtitle("Body Mass") ///  
> ytitle("Culmen Length")



scatterplot

# Linear Fit: twoway lfit y x

. twoway lfit culmen\_length\_mm body\_mass\_g, ///  
> title("Penguin Culmen Length by Body Mass") ///   
> xtitle("Body Mass") ///  
> ytitle("Culmen Length")



scatterplot