??? Introduction to data visualizations in R ???

Akanksha Garlapati

2024-09-09

Table of Contents

# Introduction to data visualizations in R

## Loading packages and datasets

library (tidyverse)

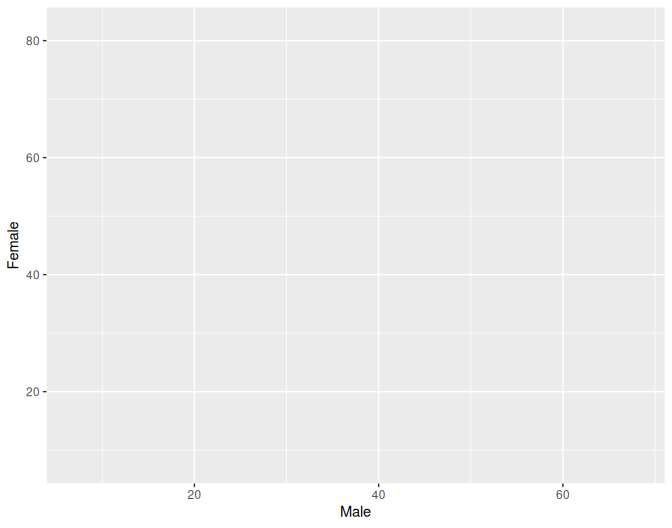
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library (modeldata)  
  
library(readxl)  
my\_dataset <- read\_excel("Animals\_and\_share \_of\_US\_adults.xlsx")  
View(my\_dataset)

## Using scatter plot

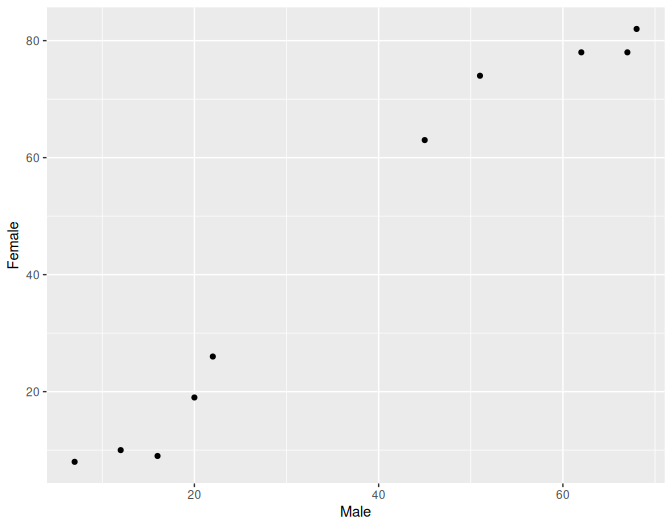
We specify dataset to be plotted in ggplot, and the variables and other aesthetics in aes

ggplot(my\_dataset, aes(x=Male,  
 y=Female))



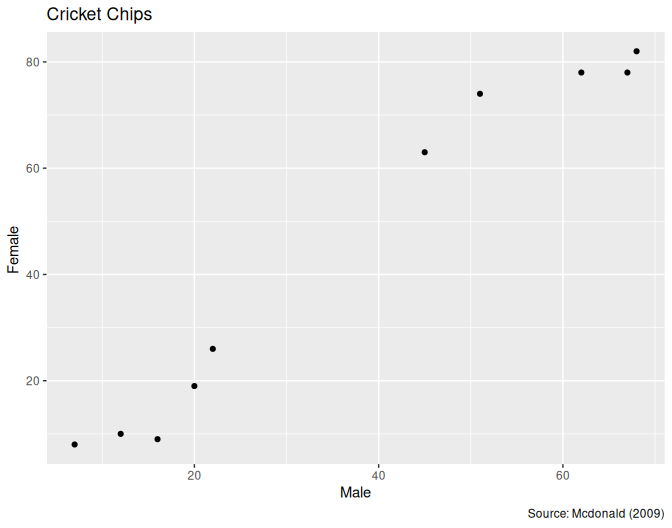
We the add the specification for the type of plot with geom\_\*

ggplot (my\_dataset, aes(x=Male,  
 y = Female ))+  
 geom\_point()



We can then add other things like labels

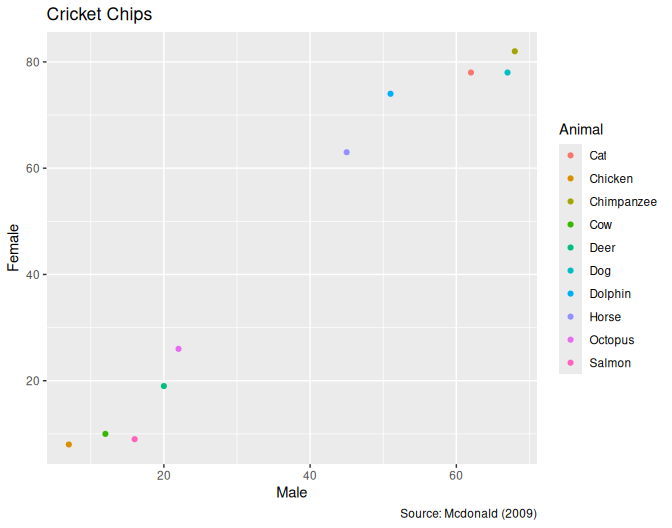
ggplot (my\_dataset, aes(x = Male,  
 y = Female ))+  
 geom\_point()+  
 labs(x="Male",  
 y="Female ",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")



### Specifying other aesthetics

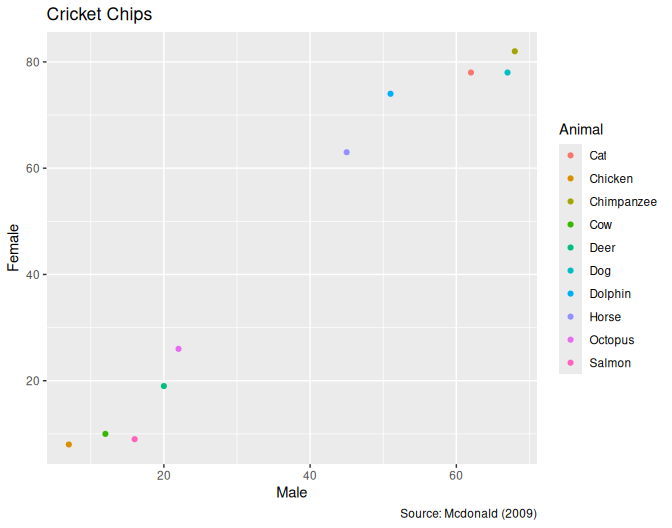
Add color with a variable

ggplot (my\_dataset, aes(x=Male,  
 y=Female,  
 color=Animal ))+  
 geom\_point()+  
 labs(x="Male",  
 y="Female",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")



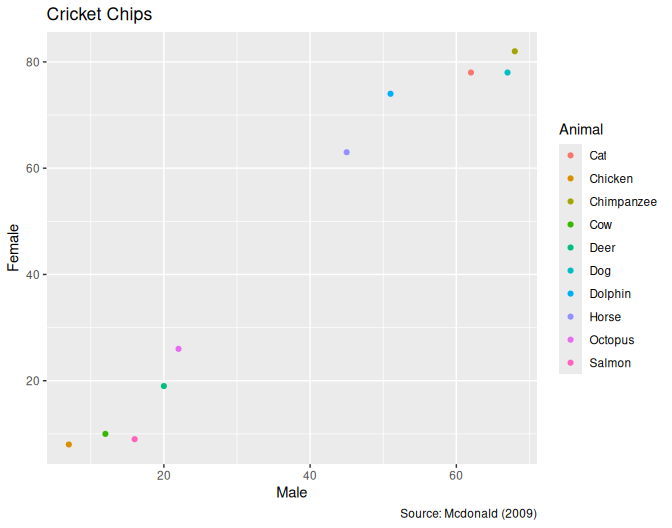
Label color labels

ggplot (my\_dataset, aes(x=Male,  
 y=Female,  
 color =Animal))+  
 geom\_point()+  
 labs(x="Male",  
 y="Female",  
 color="Animal",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")



Add color scaling

ggplot (my\_dataset, aes(x=Male,  
 y=Female,  
 color =Animal))+  
 geom\_point()+  
 labs(x="Male",  
 y="Female",  
 color="Animal",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")

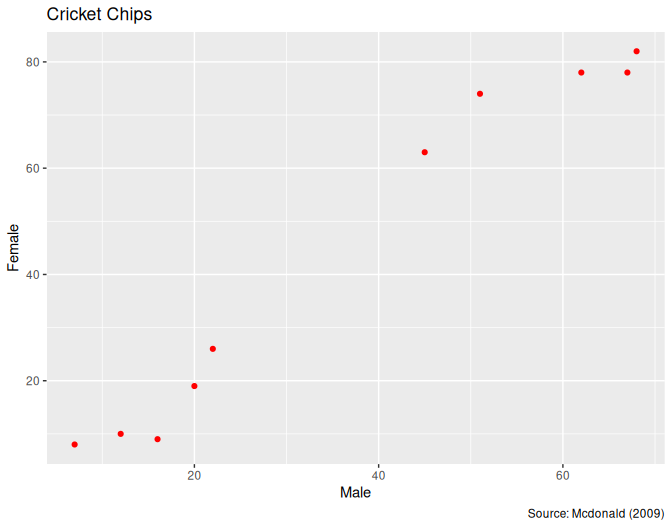


# not found: # soace\_color\_brewer(palette="Dark2")

### Modifying basic properties

Change plot color

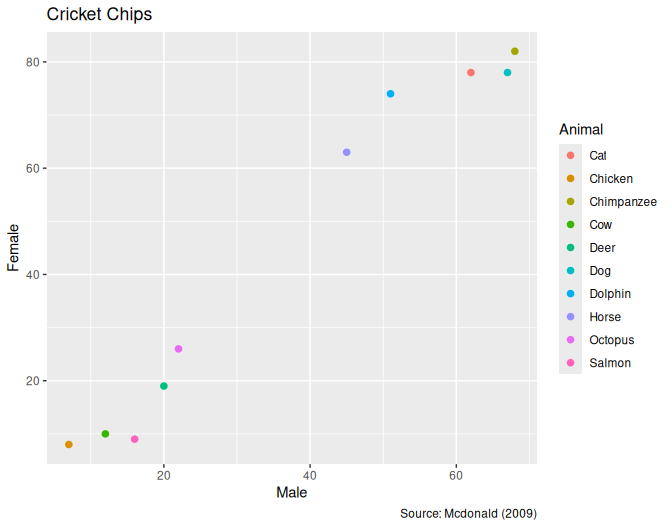
ggplot (my\_dataset, aes(x=Male,  
 y=Female,  
 color =Animal))+  
 geom\_point(color="Red")+  
 labs(x="Male",  
 y="Female",  
 color="Animal",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")



# soace\_color\_brewer(palette="Dark2")

change plot size

ggplot (my\_dataset, aes(x=Male,y=Female,color =Animal))+  
 geom\_point(size=2)+  
 labs(x="Male",  
 y="Female",  
 color="Animal",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")

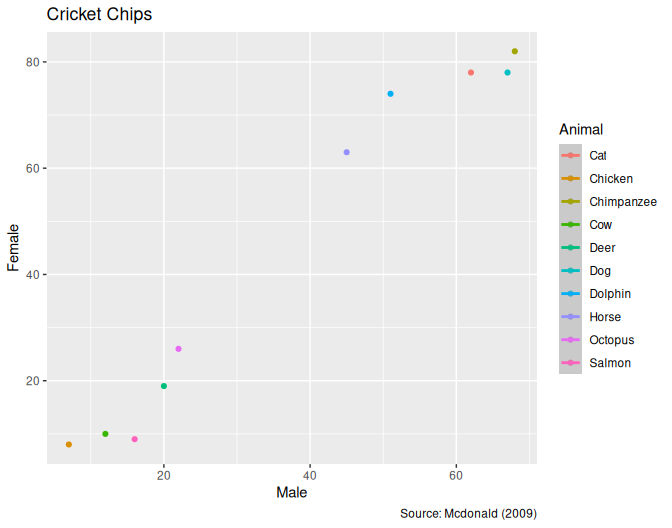


# soace\_color\_brewer(palette="Dark2")

Adding another plot layer

ggplot (my\_dataset, aes(x=Male,  
 y=Female,  
 color =Animal))+  
 geom\_point()+  
 geom\_smooth()+  
 labs(x="Male",  
 y="Female",  
 color="Animal",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

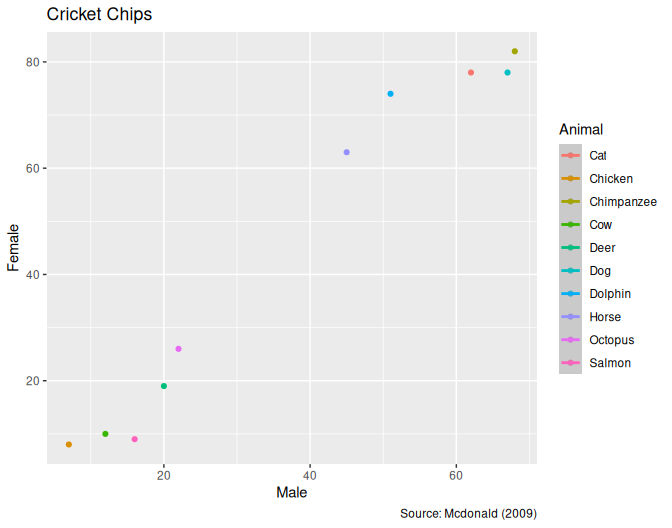


# soace\_color\_brewer(palette="Dark2")

Add geom\_smooth method

ggplot (my\_dataset, aes(x=Male,  
 y=Female,  
 color =Animal))+  
 geom\_point()+  
 geom\_smooth(method="lm")+  
 labs(x="Male",  
 y="Female",  
 color="Animal",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")

## `geom\_smooth()` using formula = 'y ~ x'

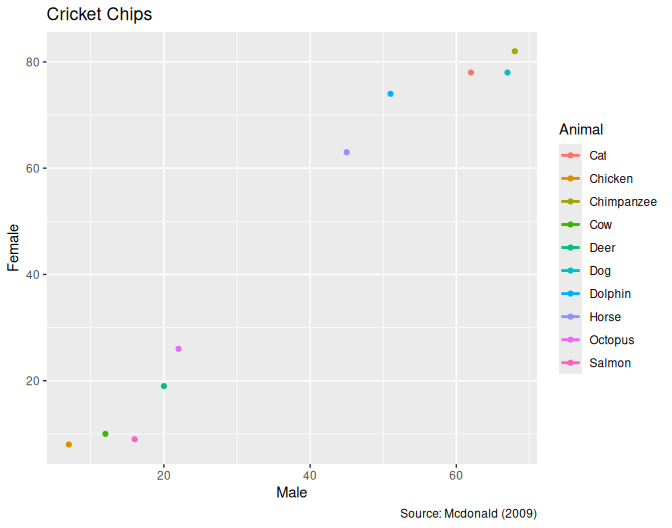


# soace\_color\_brewer(palette="Dark2")

Remove geom\_smooth error approximation

ggplot (my\_dataset, aes(x=Male,  
 y=Female,  
 color =Animal))+  
 geom\_point()+  
 geom\_smooth(method="lm",  
 se=FALSE)+  
 labs(x="Male",  
 y="Female",  
 color="Animal",  
 title="Cricket Chips",  
 caption= "Source: Mcdonald (2009)")

## `geom\_smooth()` using formula = 'y ~ x'



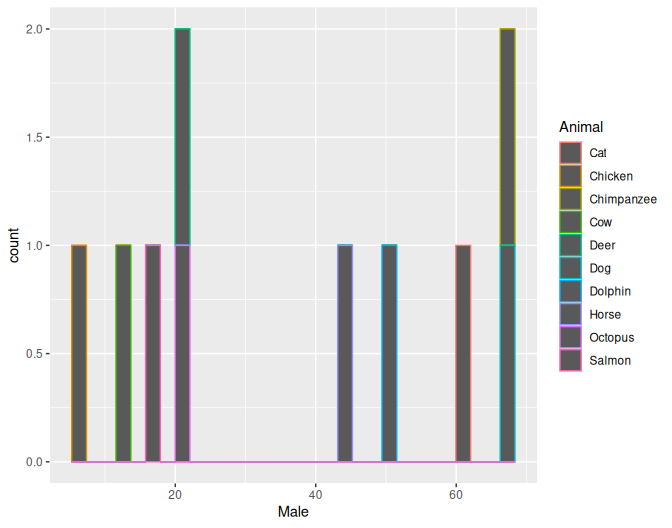
# soace\_color\_brewer(palette="Dark2")

## Other plots

### Histogram

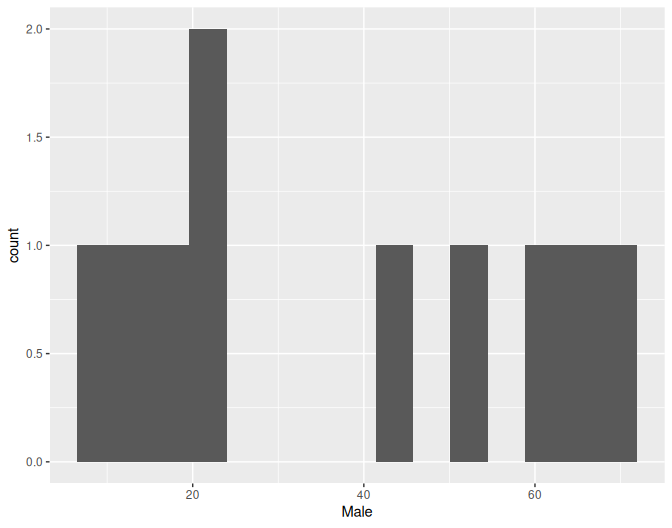
ggplot (my\_dataset, aes(x=Male,  
 color =Animal))+  
 geom\_histogram()

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



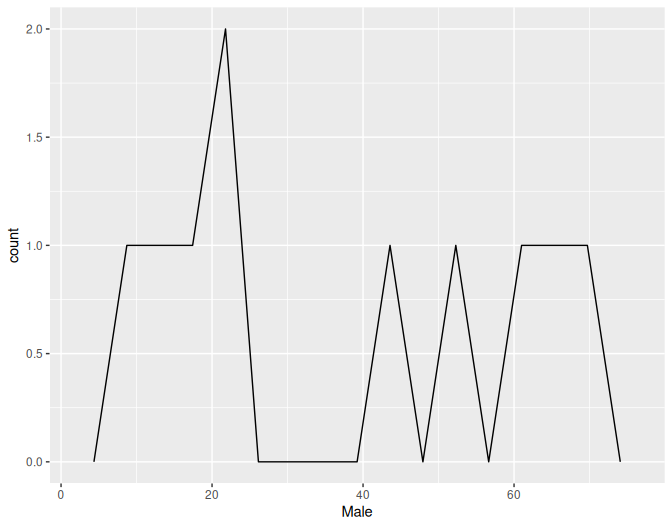
Modifying Histogram bins

ggplot (my\_dataset, aes(x=Male))+  
 geom\_histogram(bins=15)



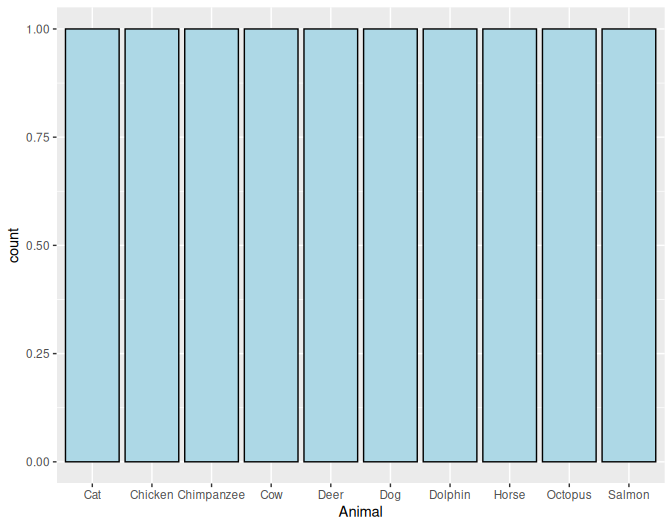
### Frequency polygon

ggplot (my\_dataset, aes(x=Male))+  
 geom\_freqpoly(bins=15)



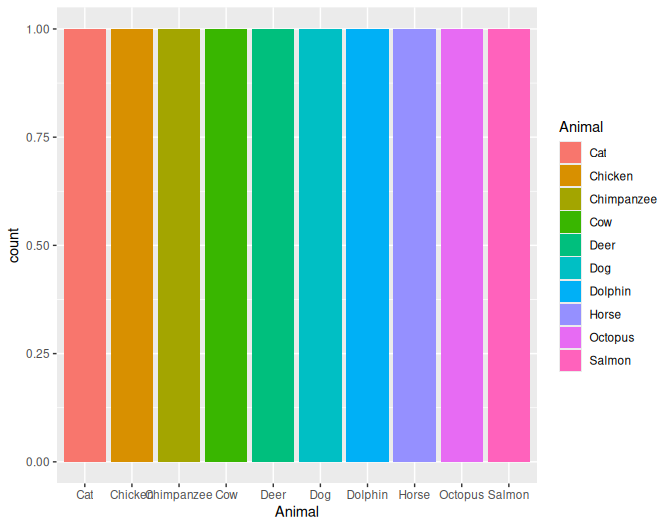
### Bars

ggplot (my\_dataset, aes(x=Animal))+  
 geom\_bar(color="black",  
 fill="lightblue")



Coloring with classes

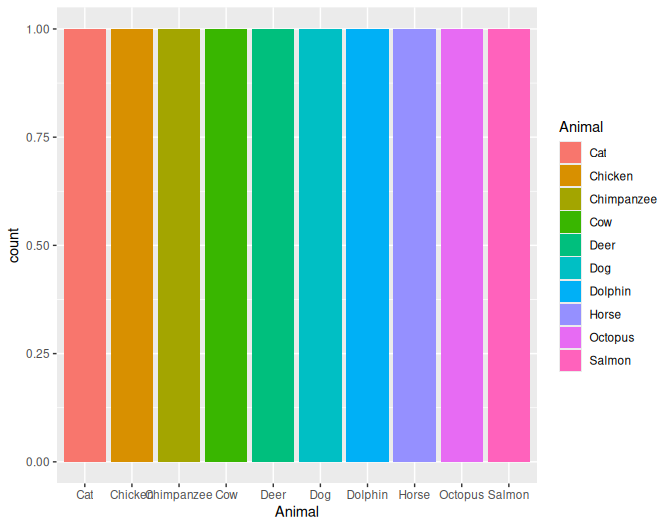
ggplot (my\_dataset, aes(x=Animal,  
 fill=Animal))+  
 geom\_bar()



# soace\_fill\_brewer(palette="Dark2")

Remove legend

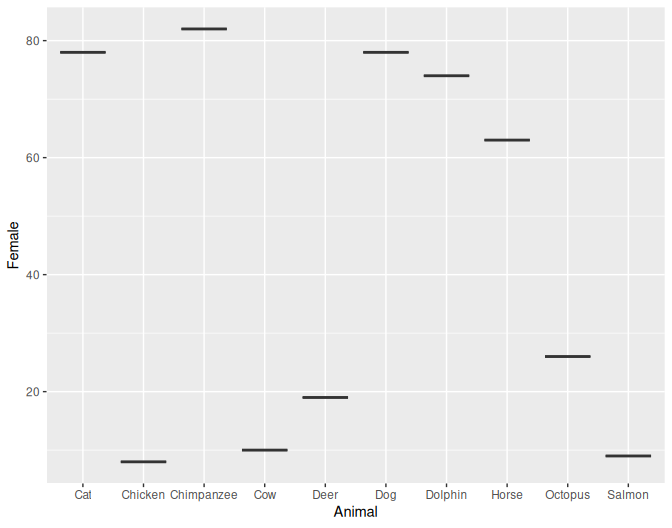
ggplot (my\_dataset, aes(x=Animal,  
 fill=Animal))+  
 geom\_bar()



# soace\_fill\_brewer(palette="Dark2")

### Box plot

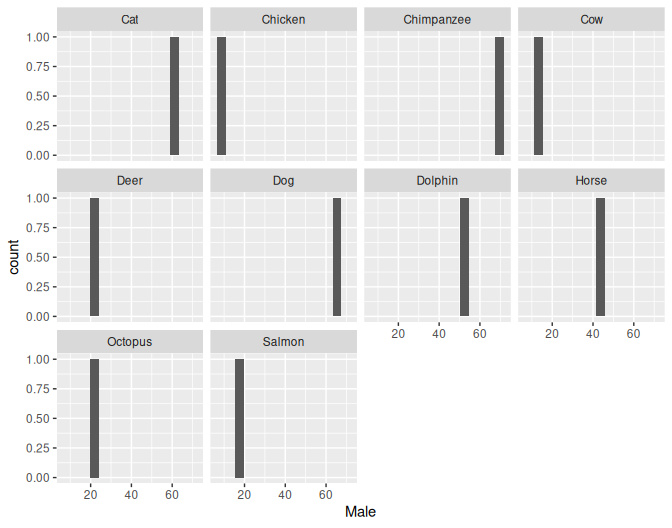
ggplot (my\_dataset, aes(x=Animal,y=Female))+  
 geom\_boxplot()



# soace\_fill\_brewer(palette="Dark2")

## Faceting

ggplot (my\_dataset, aes(x=Male))+  
 geom\_histogram(bins=15)+  
 facet\_wrap(~Animal)



# soace\_fill\_brewer(palette="Dark2")