penis project

library(ggplot2)  
library(tidyverse)

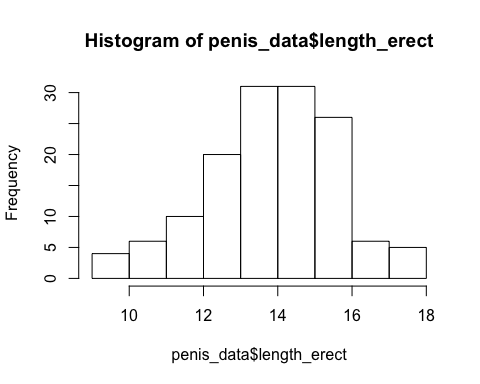
## ── Attaching packages ───────────────────────────────────────────────────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ tibble 2.1.3 ✓ dplyr 0.8.5  
## ✓ tidyr 1.0.2 ✓ stringr 1.4.0  
## ✓ readr 1.3.1 ✓ forcats 0.4.0  
## ✓ purrr 0.3.3

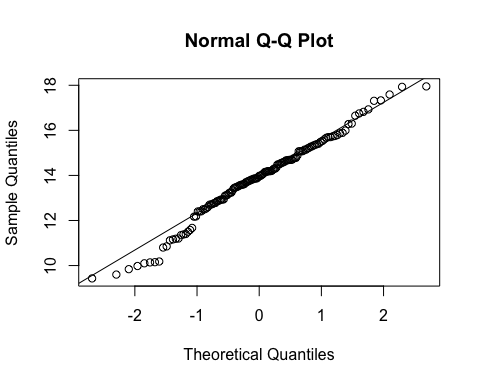
## ── Conflicts ──────────────────────────────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

# read in dataset  
penis\_data <- read.csv("/Users/Dohyun/Desktop/projects/Penis-Project/world\_penis\_dataset/penis.csv")

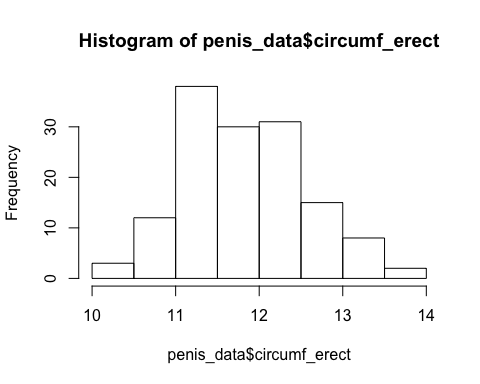
#check normality of erect length means  
  
#using a histogram  
hist(penis\_data$length\_erect)



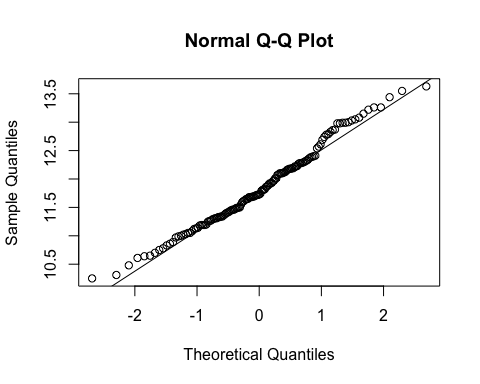
#NPP plot  
qqnorm(penis\_data$length\_erect)  
qqline(penis\_data$length\_erect)



#check normality of erect girth means  
  
#using a histogram  
hist(penis\_data$circumf\_erect)



#NPP plot  
qqnorm(penis\_data$circumf\_erect)  
qqline(penis\_data$circumf\_erect)



t.test(penis\_data$length\_erect)

##   
## One Sample t-test  
##   
## data: penis\_data$length\_erect  
## t = 91.633, df = 138, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 13.55726 14.15526  
## sample estimates:  
## mean of x   
## 13.85626

t.test(penis\_data$circumf\_erect)

##   
## One Sample t-test  
##   
## data: penis\_data$circumf\_erect  
## t = 192.43, df = 138, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 11.71941 11.96275  
## sample estimates:  
## mean of x   
## 11.84108

Confidence interval for mean erect length is 13.56-14.16 cm. Confidence interval for mean erect girth is 11.72-11.96 cm.

#check for biases  
self\_reported\_data <- filter(penis\_data, Method == "Self reported")  
measured\_data <- filter(penis\_data, Method == "Measured")  
  
t.test(self\_reported\_data$length\_erect)

##   
## One Sample t-test  
##   
## data: self\_reported\_data$length\_erect  
## t = 79.429, df = 50, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 14.33515 15.07897  
## sample estimates:  
## mean of x   
## 14.70706

t.test(self\_reported\_data$circumf\_erect)

##   
## One Sample t-test  
##   
## data: self\_reported\_data$circumf\_erect  
## t = 139.45, df = 50, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 11.90118 12.24902  
## sample estimates:  
## mean of x   
## 12.0751

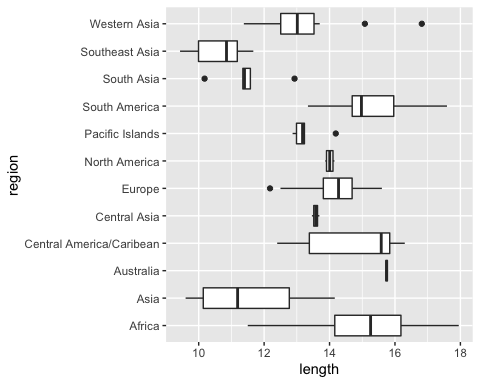
t.test(measured\_data$length\_erect)

##   
## One Sample t-test  
##   
## data: measured\_data$length\_erect  
## t = 68.323, df = 87, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 12.97443 13.75193  
## sample estimates:  
## mean of x   
## 13.36318

t.test(measured\_data$circumf\_erect)

##   
## One Sample t-test  
##   
## data: measured\_data$circumf\_erect  
## t = 146.16, df = 87, p-value < 2.2e-16  
## alternative hypothesis: true mean is not equal to 0  
## 95 percent confidence interval:  
## 11.54628 11.86463  
## sample estimates:  
## mean of x   
## 11.70545

size\_length <- length(penis\_data$length\_erect)  
size\_girth <- length(penis\_data$circumf\_erect)  
region <- penis\_data[,"Region"]  
length <- penis\_data[,"length\_erect"]  
girth <- penis\_data[,"circumf\_erect"]  
  
#boxplot of the regions  
bp <- ggplot(penis\_data, aes(x = region, y = length)) +  
 geom\_boxplot()  
bp + coord\_flip()



lm(length ~ region)

##   
## Call:  
## lm(formula = length ~ region)  
##   
## Coefficients:  
## (Intercept) regionAsia   
## 15.28136 -3.62906   
## regionAustralia regionCentral America/Caribean   
## 0.46364 -0.55836   
## regionCentral Asia regionEurope   
## -1.70636 -1.06811   
## regionNorth America regionPacific Islands   
## -1.27386 -2.01636   
## regionSouth America regionSouth Asia   
## 0.09364 -3.79336   
## regionSoutheast Asia regionWestern Asia   
## -4.66708 -2.02574