library(dplyr)

**# CALL DATASHEET BY SETTING WORKING DIRECTORY**

getwd()

setwd("C:/Users/Sophi/Documents/GitHub/TanakekeProject/RClass/UNHAS")

**#Download gapminder data first**

gapminder <- read.csv("gapminder.csv")

**#OR**

gapminder <- read.csv("C:/Users/Sophi/Documents/GitHub/TanakekeProject/RClass/UNHAS/gapminder.csv")

View(gapminder) #describe what this dataset is doing

**# GRAPHING boxplot and line DPLYR filter ----------------------------------------------------------------**

**# Boxplot gdp per capita. Then filter by country**

boxplot(lifeExp ~ country, data = gapminder)

**# DPLYR just looking at specific countries**

SEAsia <- gapminder %>%

filter(country %in% c("Indonesia", "Malaysia", "Singapore"))

boxplot(pop ~ country, data = SEAsia)

**#Look at line graph**

Indonesia <- gapminder %>%

filter(country == "Indonesia")

plot(x = Indonesia$year, y = Indonesia$lifeExp, type = 'l')

**# SHOW HOW TO EXTRACT A GRAPH**

**# GRAPHING scatterplot DPLYR group by and summarize ----------------------------------------------------------------**

plot(x = gapminder$gdpPercap, y = gapminder$lifeExp)

lifeExpectancyData <- gapminder %>%

group\_by(country) %>%

summarize(AvgLife = mean(lifeExp), Avggdp = mean(gdpPercap)) #look at the file we just made

View(lifeExpectancyData)

plot(x = lifeExpectancyData$Avggdp, y = lifeExpectancyData$AvgLife)

**# GRAPHING scatterplot DPLYR mutate ----------------------------------------------------------------**

gdpSEAsia <- SEAsia %>%

group\_by(country) %>%

summarise(AvgLife = mean(lifeExp), Avggdp = mean(gdpPercap))

barplot(height = gdpSEAsia$Avggdp, names.arg = gdpSEAsia$country)

gdpSEAsia\_IDR <- gdpSEAsia %>%

mutate(gdpIDR = Avggdp \* 15491)

barplot(height = gdpSEAsia\_IDR$gdpIDR, names.arg = gdpSEAsia\_IDR$country)

**# GRAPHING export csv file DPLYR select and rename ----------------------------------------------------------------**

**# Our boss wants population over time of each country. But they only speak indonesian**

poplasiDunia <- gapminder %>%

select(country, year, pop) %>%

rename(negara = country,

tahun = year,

poplasi = pop)

write.csv(poplasiDunia,"dataPoplasi.csv")

**#### Return to code**

**# Statistics in R ----------------------------------------------------------------**

dat <- read.csv("https://raw.githubusercontent.com/ucdavis-bioinformatics-training/2018-September-Bioinformatics-Prerequisites/master/friday/lm\_example\_data.csv")

**#You can also read csv data directly off of the internet**

**#WHAT IS THIS DATASHEET? WHAT ARE ALL OF THE VARIABLES AND TYPES OF VARIABLES**

**#Statistics to show distributions**

**# Histogram**

hist(dat$expression)

hist(dat$expression, breaks = 20)

**# Boxplots also show data distributions**

boxplot(expression ~ treatment, data = dat)

**# ANOVAS and LMs**

one.way\_ANOVA <- lm(expression ~ treatment, data = dat) #using a categorical with a continuous

summary(one.way\_ANOVA)

**# Linear Models**

linearModel <- lm(expression ~ temperature, data = dat)

summary(linearModel)

plot(x = dat$temperature, y = dat$expression)

abline(lm(expression ~ temperature, data = dat))

**#### Return to code TO COMPARE**

**#diagnostics**

plot(linearModel)

# DOES ANYONE HAVE ANY OTHER SPECIFIC STATISTICAL TESTS THEY WOULD LIKE TO RUN

**#### Return to code**

NEXT LESSON: DATES

DATES IN R

library(lubridate)

library(dplyr)

today()

now()

myDate <- as.Date("2015-10-19 10:15")

str(myDate)

timeDate <- as.POSIXct("2015-10-19 10:15")

str(timeDate)

timeDatelt<- as.POSIXlt("2015-10-19 10:15")

str(timeDatelt)

dateInfo <- as.POSIXct("January 30, 25") #Doesn't work

dateInfo <- as.POSIXct("January 30, 25 10:15 PM", format = "%B %d, %y %I:%M %p")

str(dateInfo)

**#Extracting information from**

year(dateInfo)

month(dateInfo)

day(dateInfo)

hour(dateInfo)

minute(dateInfo)

**# ACTIVITY: READ IN YOUR CLEANED DATA FROM THE DATA HYGIENE LESSON AND READ IN THE DATE COLUMN AS A DATE.**

**# MAKE A SCATTER PLOT OF THE WEIGHTS OVER TIME**

**# ACTIVITY: DOWNLOAD THE DATES DATASET. EACH GROUP WILL PICK A DIFFERENT COLUMN AND FIX THE DATES AND GRAPH A TIME SERIES**

dateData <- read.csv("C:/Users/Sophi/Documents/GitHub/TanakekeProject/RClass/UNHAS/dateTime.csv")

**# Try a time series alone**

plot(x = dateData$date3, y = dateData$sales, type = 'l')

class(dateData$date3)

**#Must change the datatype**

dateData$date <- as.POSIXct(dateData$date3, format = "%A, %B %d, %Y")

class(dateData$date)

plot(x = dateData$date3, y = dateData$sales, type = 'l')

**# ACTIVITY: USE DPLYR AND GRAPHING TO EXTRACT AND SHOW TH EMONLTY AVERAGE SALES ACROSS ALL OF THE YEARS**

monthlyAverage <- dateData %>%

mutate(month = month(date)) %>%

group\_by(month) %>%

summarize(mean = mean(sales))

monthlyAverage

barplot(height = monthlyAverage$mean, names.arg = monthlyAverage$month)