StudentSurvey

Kyle Ramirez

12/26/2021

```
{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)
```

Use the apply function on a variable in your dataset

importing libraries

library(ggplot2) theme_set(theme_minimal())

Set the working directory to the root of your DSC 520 directory

setwd("/Users/Kyle/Documents/GitHub/KR/Ramirez_Kyle_DSC510/dsc520/data") surveyData_df <- read.csv('student-survey.csv', stringsAsFactors = FALSE)

apply(X=surveyData_df, MARGIN=2, FUN=mean)

Use the aggregate function on a variable in your dataset

aggregate(TimeReading~Happiness, surveyData_df, mean)

Use the plyr function on a variable in your dataset – more specifically, I want to see you split some data, perform a modification to the data, and then bring it back together

library(dplyr) Female <- filter(surveyData_df, Gender == 1) Male <- filter(surveyData_df, Gender == 0)

Check distributions of the data

ggplot(surveyData_df, aes(x=Gender, y=Happiness)) + geom_point() + ggtitle('Male & Female Happiness') + xlab('Males are 0, Females are 1') + ylab('Happiness Score')

Identify if there are any outliers

A far outlier for males below 50

Create at least 2 new variables

Female_df <- subset(surveyData_df, Gender == 1) Male_df <- subset(surveyData_df, Gender == 0)