Exam 2 PDF

Madison Pape

6/28/2021

## Question 1 - Clear the environment

rm(list = ls())

## Question 2 - Load and name exam data

library(readxl)  
college\_scorecard <- read\_excel("2021\_exam2\_data - college scorecard.xlsx")  
attach(college\_scorecard)

## Question 3 - Providing summary statistcs

summary(college\_scorecard)

## Question 4 - Creating smaller dataset for 2014-2015 graduates of 4+ year colleges and universities located in TX and LA

library(tidyverse)  
small\_scorecard\_year <-  
 college\_scorecard %>%  
 dplyr::filter(year=="2014" | year=="2015")   
   
small\_scorecard <-  
 small\_scorecard\_year %>%  
 dplyr::filter(state\_abbr=="LA" | state\_abbr=="TX")

## Question 5 - Collapse the small\_scorecard data frame to get both the average of number people working who graduated from universities in Texas and Lousiana; and the total number of people working who graduated from universities in Texas and Lousiana. Call your resulting data frame “even\_smaller\_scorecard”

tx\_total <- sum(small\_scorecard[small\_scorecard$state\_abbr == “TX”]$count\_working)  
la\_total <- sum(small\_scorecard[small\_scorecard$state\_abbr == “LA”]$count\_working)  
  
tx\_count <- sum(small\_scorecard$state\_abbr == “TX”)  
la\_count <- sum(small\_scorecard$state\_abbr == “LA”)  
  
even\_smaller\_scorecard <- data.frame(“State” = c(“Texas”,”Louisiana”), “Average” = c(tx\_total/tx\_count, la\_total/la\_count), “Total” = c(tx\_total, la\_total))

## Question 6 - Create a bar graph of percentage of people working

total\_allwork <- sum(count\_working) + sum(count\_not\_working)  
percent\_working <- sum(count\_working) / total\_allwork  
barplot(percent\_working)

## Question 7 -

## Question 8 - Load data frame and call it “Avocados”

library(readxl)  
avocados <- read\_excel("2021\_exam2\_data - avocados.xlsx")  
attach(avocados)

## Question 9 - Create a new variable “year” that only captures year sold

library(tidyverse)  
year <- lubridate::year(avocados$date)

## Question 10 -

library(WDI)  
average\_price\_adjusted <-

## Question 11 - collapsing the data

collapsed\_avocados <-  
 avocados %>%  
 group\_by(average\_price\_adjusted) %>%  
 head(collapsed\_avocados)

## Question 12 -

## Question 13 -

## Question 14 - load the training dataset and call it training

library(readxl)  
training <- read\_excel("2021\_exam2\_data - titanic.xlsx")  
View(training)

## Question 15 -

## Question 16 - load the Titanic dataset

library(readxl)  
titanic <- read\_excel("2021\_exam2\_data - titanic.xlsx")  
View(titanic)  
attatch(titanic)

## Question 17 - provide summary statistics for the titanic data

summary(titanic)

## Question 18 - create a cross-tab for survivorship by gender

gender\_survival <- table(titanic$female, titanic$survived)  
View(gender\_survival)

Knowing that variable 1 in our table governs gender, with 1 being femmale and 0 being male, and variable 2 in our table governs survivorship, with 0 meaning did not survive and 1 meaning did survive, we can determine that more men than women did die on the titanic. More men died than women overall but there were also more men who survived as there were more men than women on the boat in general.

## Question 19 - frequency table for first\_class

ifelse(titanic$class=="1", "nothing")  
titanic %>%  
 table(titanic)

## Bonus - “My Heart Will Go On” by Celine Dion