p8105\_hw1\_yk2960

Youyuan(Keviant) Kong

2021/9/23

##Problem 1

random\_sample<-rnorm(10)  
###a random sample of size 10 from a standard Normal distribution  
logical\_vector<-random\_sample>0  
##a logical vector indicating whether elements of the sample are greater than 0  
char\_vector<-rep("10",10)  
##a character vector of length 10  
factor\_vector<-c("1","2","3","1","2","3","1","2","3","1")  
factor\_vector<-as.factor(factor\_vector)  
##a factor vector of length 10, with 3 different factor “levels”  
dataFrame\_problem1<-data.frame(random\_sample,logical\_vector,char\_vector,factor\_vector)  
##take the mean of a variable in a dataframe  
library(tidyverse)

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.4 v dplyr 1.0.7  
## v tidyr 1.1.3 v stringr 1.4.0  
## v readr 2.0.1 v forcats 0.5.1

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

mean(pull(dataFrame\_problem1,1))

## [1] -0.08570232

##numeric can get mean  
mean(pull(dataFrame\_problem1,2))

## [1] 0.4

##logical can get mean  
mean(pull(dataFrame\_problem1,3))

## Warning in mean.default(pull(dataFrame\_problem1, 3)): 参数不是数值也不是逻辑值：  
## 回覆NA

## [1] NA

mean(as.numeric(pull(dataFrame\_problem1,3)))

## [1] 10

##char cannot get mean  
mean(pull(dataFrame\_problem1,4))

## Warning in mean.default(pull(dataFrame\_problem1, 4)): 参数不是数值也不是逻辑值：  
## 回覆NA

## [1] NA

mean(as.numeric(pull(dataFrame\_problem1,4)))

## [1] 1.9

##char cannot get mean

##Problem 2

##install.packages("palmerpenguins")  
library(palmerpenguins)  
data("penguins", package = "palmerpenguins")  
summary(penguins\_raw)

## studyName Sample Number Species Region   
## Length:344 Min. : 1.00 Length:344 Length:344   
## Class :character 1st Qu.: 29.00 Class :character Class :character   
## Mode :character Median : 58.00 Mode :character Mode :character   
## Mean : 63.15   
## 3rd Qu.: 95.25   
## Max. :152.00   
##   
## Island Stage Individual ID Clutch Completion   
## Length:344 Length:344 Length:344 Length:344   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Date Egg Culmen Length (mm) Culmen Depth (mm) Flipper Length (mm)  
## Min. :2007-11-09 Min. :32.10 Min. :13.10 Min. :172.0   
## 1st Qu.:2007-11-28 1st Qu.:39.23 1st Qu.:15.60 1st Qu.:190.0   
## Median :2008-11-09 Median :44.45 Median :17.30 Median :197.0   
## Mean :2008-11-27 Mean :43.92 Mean :17.15 Mean :200.9   
## 3rd Qu.:2009-11-16 3rd Qu.:48.50 3rd Qu.:18.70 3rd Qu.:213.0   
## Max. :2009-12-01 Max. :59.60 Max. :21.50 Max. :231.0   
## NA's :2 NA's :2 NA's :2   
## Body Mass (g) Sex Delta 15 N (o/oo) Delta 13 C (o/oo)  
## Min. :2700 Length:344 Min. : 7.632 Min. :-27.02   
## 1st Qu.:3550 Class :character 1st Qu.: 8.300 1st Qu.:-26.32   
## Median :4050 Mode :character Median : 8.652 Median :-25.83   
## Mean :4202 Mean : 8.733 Mean :-25.69   
## 3rd Qu.:4750 3rd Qu.: 9.172 3rd Qu.:-25.06   
## Max. :6300 Max. :10.025 Max. :-23.79   
## NA's :2 NA's :14 NA's :13   
## Comments   
## Length:344   
## Class :character   
## Mode :character   
##   
##   
##   
##

str(penguins\_raw)

## tibble [344 x 17] (S3: tbl\_df/tbl/data.frame)  
## $ studyName : chr [1:344] "PAL0708" "PAL0708" "PAL0708" "PAL0708" ...  
## $ Sample Number : num [1:344] 1 2 3 4 5 6 7 8 9 10 ...  
## $ Species : chr [1:344] "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pygoscelis adeliae)" "Adelie Penguin (Pygoscelis adeliae)" ...  
## $ Region : chr [1:344] "Anvers" "Anvers" "Anvers" "Anvers" ...  
## $ Island : chr [1:344] "Torgersen" "Torgersen" "Torgersen" "Torgersen" ...  
## $ Stage : chr [1:344] "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" "Adult, 1 Egg Stage" ...  
## $ Individual ID : chr [1:344] "N1A1" "N1A2" "N2A1" "N2A2" ...  
## $ Clutch Completion : chr [1:344] "Yes" "Yes" "Yes" "Yes" ...  
## $ Date Egg : Date[1:344], format: "2007-11-11" "2007-11-11" ...  
## $ Culmen Length (mm) : num [1:344] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42 ...  
## $ Culmen Depth (mm) : num [1:344] 18.7 17.4 18 NA 19.3 20.6 17.8 19.6 18.1 20.2 ...  
## $ Flipper Length (mm): num [1:344] 181 186 195 NA 193 190 181 195 193 190 ...  
## $ Body Mass (g) : num [1:344] 3750 3800 3250 NA 3450 ...  
## $ Sex : chr [1:344] "MALE" "FEMALE" "FEMALE" NA ...  
## $ Delta 15 N (o/oo) : num [1:344] NA 8.95 8.37 NA 8.77 ...  
## $ Delta 13 C (o/oo) : num [1:344] NA -24.7 -25.3 NA -25.3 ...  
## $ Comments : chr [1:344] "Not enough blood for isotopes." NA NA "Adult not sampled." ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. studyName = col\_character(),  
## .. `Sample Number` = col\_double(),  
## .. Species = col\_character(),  
## .. Region = col\_character(),  
## .. Island = col\_character(),  
## .. Stage = col\_character(),  
## .. `Individual ID` = col\_character(),  
## .. `Clutch Completion` = col\_character(),  
## .. `Date Egg` = col\_date(format = ""),  
## .. `Culmen Length (mm)` = col\_double(),  
## .. `Culmen Depth (mm)` = col\_double(),  
## .. `Flipper Length (mm)` = col\_double(),  
## .. `Body Mass (g)` = col\_double(),  
## .. Sex = col\_character(),  
## .. `Delta 15 N (o/oo)` = col\_double(),  
## .. `Delta 13 C (o/oo)` = col\_double(),  
## .. Comments = col\_character()  
## .. )

nrow(penguins\_raw)

## [1] 344

ncol(penguins\_raw)

## [1] 17

complete\_penguins\_raw<-penguins\_raw[complete.cases(penguins\_raw$`Flipper Length (mm)`),]  
mean(as.numeric(complete\_penguins\_raw$`Flipper Length (mm)`))

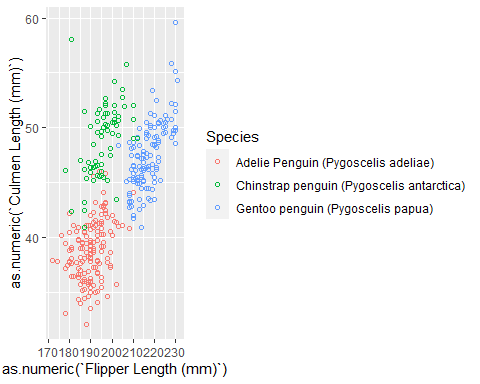
## [1] 200.9152

penguins\_raw$`Culmen Length (mm)`

## [1] 39.1 39.5 40.3 NA 36.7 39.3 38.9 39.2 34.1 42.0 37.8 37.8 41.1 38.6 34.6  
## [16] 36.6 38.7 42.5 34.4 46.0 37.8 37.7 35.9 38.2 38.8 35.3 40.6 40.5 37.9 40.5  
## [31] 39.5 37.2 39.5 40.9 36.4 39.2 38.8 42.2 37.6 39.8 36.5 40.8 36.0 44.1 37.0  
## [46] 39.6 41.1 37.5 36.0 42.3 39.6 40.1 35.0 42.0 34.5 41.4 39.0 40.6 36.5 37.6  
## [61] 35.7 41.3 37.6 41.1 36.4 41.6 35.5 41.1 35.9 41.8 33.5 39.7 39.6 45.8 35.5  
## [76] 42.8 40.9 37.2 36.2 42.1 34.6 42.9 36.7 35.1 37.3 41.3 36.3 36.9 38.3 38.9  
## [91] 35.7 41.1 34.0 39.6 36.2 40.8 38.1 40.3 33.1 43.2 35.0 41.0 37.7 37.8 37.9  
## [106] 39.7 38.6 38.2 38.1 43.2 38.1 45.6 39.7 42.2 39.6 42.7 38.6 37.3 35.7 41.1  
## [121] 36.2 37.7 40.2 41.4 35.2 40.6 38.8 41.5 39.0 44.1 38.5 43.1 36.8 37.5 38.1  
## [136] 41.1 35.6 40.2 37.0 39.7 40.2 40.6 32.1 40.7 37.3 39.0 39.2 36.6 36.0 37.8  
## [151] 36.0 41.5 46.1 50.0 48.7 50.0 47.6 46.5 45.4 46.7 43.3 46.8 40.9 49.0 45.5  
## [166] 48.4 45.8 49.3 42.0 49.2 46.2 48.7 50.2 45.1 46.5 46.3 42.9 46.1 44.5 47.8  
## [181] 48.2 50.0 47.3 42.8 45.1 59.6 49.1 48.4 42.6 44.4 44.0 48.7 42.7 49.6 45.3  
## [196] 49.6 50.5 43.6 45.5 50.5 44.9 45.2 46.6 48.5 45.1 50.1 46.5 45.0 43.8 45.5  
## [211] 43.2 50.4 45.3 46.2 45.7 54.3 45.8 49.8 46.2 49.5 43.5 50.7 47.7 46.4 48.2  
## [226] 46.5 46.4 48.6 47.5 51.1 45.2 45.2 49.1 52.5 47.4 50.0 44.9 50.8 43.4 51.3  
## [241] 47.5 52.1 47.5 52.2 45.5 49.5 44.5 50.8 49.4 46.9 48.4 51.1 48.5 55.9 47.2  
## [256] 49.1 47.3 46.8 41.7 53.4 43.3 48.1 50.5 49.8 43.5 51.5 46.2 55.1 44.5 48.8  
## [271] 47.2 NA 46.8 50.4 45.2 49.9 46.5 50.0 51.3 45.4 52.7 45.2 46.1 51.3 46.0  
## [286] 51.3 46.6 51.7 47.0 52.0 45.9 50.5 50.3 58.0 46.4 49.2 42.4 48.5 43.2 50.6  
## [301] 46.7 52.0 50.5 49.5 46.4 52.8 40.9 54.2 42.5 51.0 49.7 47.5 47.6 52.0 46.9  
## [316] 53.5 49.0 46.2 50.9 45.5 50.9 50.8 50.1 49.0 51.5 49.8 48.1 51.4 45.7 50.7  
## [331] 42.5 52.2 45.2 49.3 50.2 45.6 51.9 46.8 45.7 55.8 43.5 49.6 50.8 50.2

library(ggplot2)  
plot<-ggplot(penguins\_raw, aes(x=as.numeric(`Flipper Length (mm)`), y=as.numeric(`Culmen Length (mm)`),color=Species))+ geom\_point(shape=1)   
plot

## Warning: Removed 2 rows containing missing values (geom\_point).



ggsave("Problem2 penguins' data plot",device="pdf",plot)

## Saving 5 x 4 in image

## Warning: Removed 2 rows containing missing values (geom\_point).