Assignment Week 5

Sara Monedero

2022-06-27

library (readr)  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.1 ──

## ✔ ggplot2 3.3.6 ✔ dplyr 1.0.9  
## ✔ tibble 3.1.7 ✔ stringr 1.4.0  
## ✔ tidyr 1.2.0 ✔ forcats 0.5.1  
## ✔ purrr 0.3.4

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

library(knitr)  
library(bslib)

##   
## Attaching package: 'bslib'

## The following object is masked from 'package:utils':  
##   
## page

library(ggplot2)

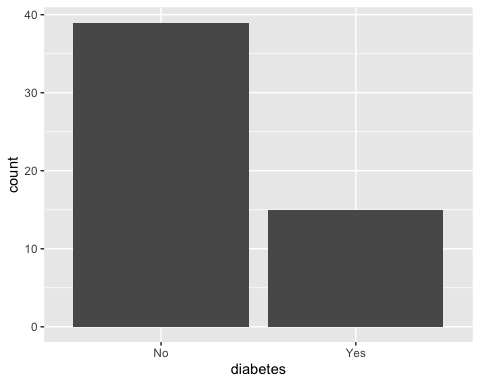
#DATA READING  
url <- "https://raw.githubusercontent.com/fivethirtyeight/data/master/nutrition-studies/raw\_anonymized\_data.csv"  
food <- read\_csv(url)

## Rows: 54 Columns: 1093  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (26): cancer, diabetes, heart\_disease, belly, ever\_smoked, currently\_s...  
## dbl (1067): ID, BREAKFASTSANDWICHFREQ, BREAKFASTSANDWICHQUAN, EGGSFREQ, EGGS...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

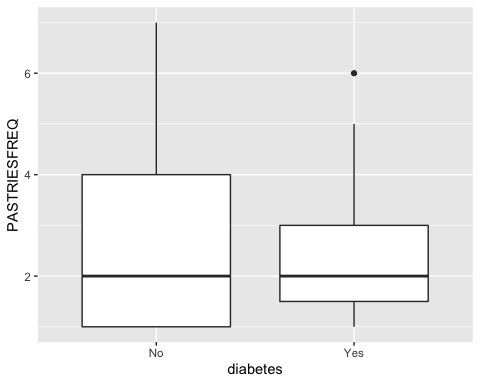
#DATA CLEANING  
diabetes <- select(food, "diabetes", "ever\_smoked","PANCAKEFREQ", "PANCAKEQUAN", "PASTRIESFREQ", "PASTRIESQUAN", "BISCUITFREQ", "BISCUITQUAN")  
  
as.data.frame(diabetes)

## diabetes ever\_smoked PANCAKEFREQ PANCAKEQUAN PASTRIESFREQ PASTRIESQUAN  
## 1 No Yes 1 1 1 1  
## 2 Yes Yes 1 1 1 1  
## 3 Yes No 3 4 6 3  
## 4 No No 4 2 5 2  
## 5 No No 2 2 4 1  
## 6 No Yes 2 2 2 2  
## 7 No No 2 3 2 2  
## 8 No No 2 2 3 2  
## 9 Yes Yes 1 1 3 2  
## 10 No No 1 1 1 1  
## 11 No Yes 3 3 4 2  
## 12 Yes No 2 3 1 1  
## 13 No No 3 2 4 2  
## 14 Yes No 3 2 2 1  
## 15 No No 1 1 4 3  
## 16 No No 3 2 3 2  
## 17 No Yes 2 2 3 2  
## 18 No Yes 2 3 4 3  
## 19 Yes Yes 1 1 1 1  
## 20 No No 4 2 5 2  
## 21 Yes No 2 3 4 2  
## 22 No Yes 4 4 1 1  
## 23 No No 3 1 2 2  
## 24 No Yes 1 1 1 1  
## 25 Yes No 3 3 3 2  
## 26 No Yes 4 2 7 2  
## 27 No No 3 3 4 2  
## 28 No No 1 1 1 1  
## 29 No No 1 1 1 1  
## 30 No No 2 2 2 2  
## 31 Yes Yes 4 1 2 2  
## 32 No Yes 4 2 2 1  
## 33 No No 4 2 4 2  
## 34 Yes No 2 3 2 3  
## 35 Yes No 3 1 2 2  
## 36 No No 4 2 2 3  
## 37 No No 2 3 2 2  
## 38 Yes No 2 4 2 2  
## 39 No No 4 3 4 3  
## 40 No No 2 2 1 1  
## 41 No Yes 1 1 1 1  
## 42 No No 2 2 1 1  
## 43 Yes No 3 2 2 3  
## 44 No No 2 1 1 1  
## 45 No No 3 2 4 2  
## 46 No Yes 2 1 2 2  
## 47 No No 3 3 3 2  
## 48 Yes No 1 1 1 1  
## 49 No No 4 1 5 2  
## 50 No No 2 3 2 3  
## 51 Yes No 1 1 5 2  
## 52 No No 3 2 3 2  
## 53 No Yes 1 1 1 1  
## 54 No No 2 2 5 3  
## BISCUITFREQ BISCUITQUAN  
## 1 2 2  
## 2 1 1  
## 3 2 4  
## 4 2 2  
## 5 1 1  
## 6 2 2  
## 7 2 2  
## 8 1 1  
## 9 1 1  
## 10 2 2  
## 11 2 2  
## 12 1 1  
## 13 3 3  
## 14 2 3  
## 15 1 1  
## 16 4 3  
## 17 3 3  
## 18 7 3  
## 19 1 1  
## 20 2 3  
## 21 4 3  
## 22 1 1  
## 23 3 3  
## 24 1 1  
## 25 1 1  
## 26 2 1  
## 27 2 1  
## 28 1 1  
## 29 1 1  
## 30 1 1  
## 31 4 2  
## 32 2 1  
## 33 2 2  
## 34 2 2  
## 35 2 3  
## 36 2 2  
## 37 1 1  
## 38 7 4  
## 39 1 1  
## 40 1 1  
## 41 1 1  
## 42 2 1  
## 43 4 2  
## 44 1 1  
## 45 2 1  
## 46 1 1  
## 47 1 1  
## 48 1 1  
## 49 1 1  
## 50 1 1  
## 51 1 1  
## 52 2 2  
## 53 2 3  
## 54 1 1

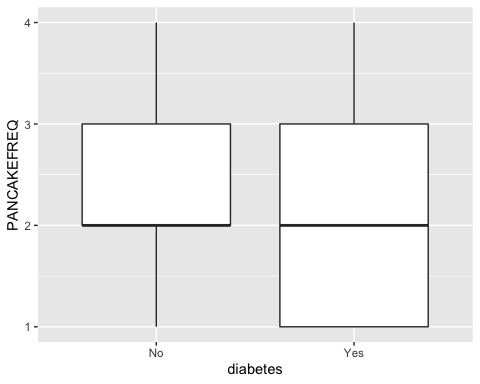
#Plot 1: Diabetes Distribution  
ggplot(data = diabetes) +  
 geom\_bar(mapping = aes(x = diabetes))



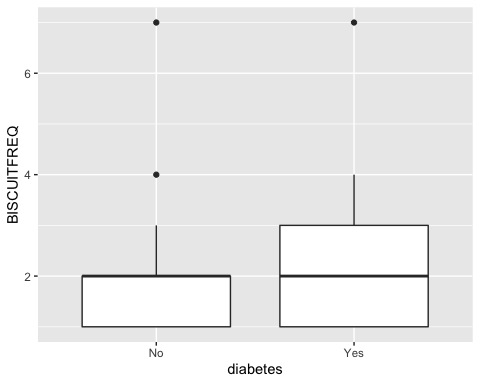
#Plot 2 - Types of sugar foods consumed by people with and without diabetes (by freq)  
ggplot(data = diabetes) +  
 geom\_boxplot(mapping = aes(x = diabetes, y = PASTRIESFREQ))



ggplot(data = diabetes) +  
 geom\_boxplot(mapping = aes(x = diabetes, y = PANCAKEFREQ))



ggplot(data = diabetes) +  
 geom\_boxplot(mapping = aes(x = diabetes, y = BISCUITFREQ))



#Plot 3 - Do freq and quantity positivately related?   
ggplot(data = diabetes) +  
 geom\_smooth(mapping = aes(x = PASTRIESQUAN, y = PASTRIESFREQ))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at 0.99

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 1.01

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 0

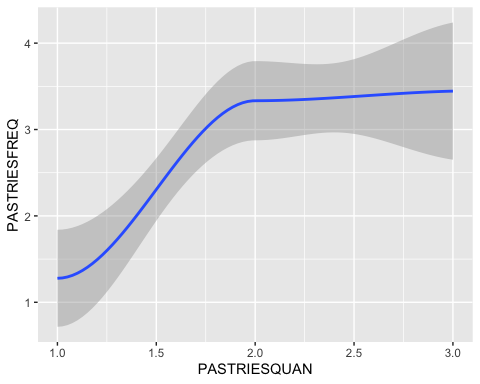
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : There are other near singularities as well. 4.0401

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at  
## 0.99

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 1.01

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition  
## number 0

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : There are other near  
## singularities as well. 4.0401



ggplot(data = diabetes) +  
 geom\_smooth(mapping = aes(x = PANCAKEQUAN, y = PANCAKEFREQ))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at 0.985

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 2.015

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 4.1466e-16

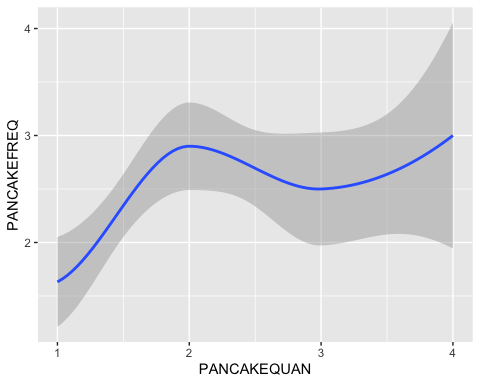
## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : There are other near singularities as well. 4

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at  
## 0.985

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius  
## 2.015

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition  
## number 4.1466e-16

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : There are other near  
## singularities as well. 4



ggplot(data = diabetes) +  
 geom\_smooth(mapping = aes(x = BISCUITQUAN, y = BISCUITFREQ))

## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : pseudoinverse used at 0.985

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : neighborhood radius 1.015

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : reciprocal condition number 0

## Warning in simpleLoess(y, x, w, span, degree = degree, parametric =  
## parametric, : There are other near singularities as well. 1

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at  
## 0.985

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius  
## 1.015

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition  
## number 0

## Warning in predLoess(object$y, object$x, newx = if  
## (is.null(newdata)) object$x else if (is.data.frame(newdata))  
## as.matrix(model.frame(delete.response(terms(object)), : There are other near  
## singularities as well. 1

