Tarea 365

Primero se carga la base de datos para tener un dataframe.

sanguchez <- read.csv2("/Users/gabrielengel/Downloads/sanguchez.csv")  
library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.3 ✓ purrr 0.3.4  
## ✓ tibble 3.1.0 ✓ dplyr 1.0.5  
## ✓ tidyr 1.1.3 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.1

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

library(ggplot2)  
library(dplyr)

Luego comienza el proceso de limpiar los datos. Lo primero que hice fue eliminar filas donde faltaba información clave.

sanguchez.sinNA <- sanguchez[complete.cases(sanguchez), ]

Después cree una nueva columna con el ID de cada sandwich. Mi primera intuición fue usar el nombre del local como el ID, pero al darme cuenta que existen algunos locales con múltiples sandwiches tuve que implementar otra columna.

vec <- c(1:402)  
sanguchez.sinNA.conID <- sanguchez.sinNA  
sanguchez.sinNA.conID$ID <- vec

Luego se omiten las columnas que no son significativas para la investigación, tales como la dirección y el link.

sanguchez.sinNA.conID.1 <- sanguchez.sinNA.conID[c(2, 4:8)]

Después, para tener un dataframe mas ordenado, se separan los ingredientes en distintas columnas para poder analizar de mejor manera.

library(stringr)  
ingredientes<- str\_split\_fixed(sanguchez.sinNA.conID.1$Ingredientes,",", 12)

Luego creamos un dataframe de los datos sin los ingredientes y la unimos al nuevo dataframe.

sangus <- sanguchez.sinNA.conID.1[c(1:2, 4:5)]  
  
ing\_cbind <- cbind(sangus, ingredientes)  
ing\_sep<- bind\_cols(sanguchez.sinNA.conID.1, ingredientes)

## New names:  
## \* NA -> ...7

ing\_cbind\_precios\_correctos <- ing\_cbind

El siguiente paso para limpiar los datos es convertir todos los precios a pesos chilenos.

ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "12 NZD"] <- "$6.051"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "12 USD"] <- "$8.605"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "17 AUD"] <- "$9.278"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == " 11 USD"] <- "$7.888"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos %in% c("15 AUD")] <- "$8.186"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "18 USD"] <- "$12.907"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "$4.200 aprox."] <- "$4.200"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "22 USD"] <- "$15.776"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "8 USD"] <- "$5.736"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "$2.000 (2 por $3.600)"] <- "$2.000"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "18 soles (Aprox $3.600)"] <- "$3.600"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "$3500 sandwich y combo $5.500"] <- "$3.500"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "15.5 AUD"] <- "$8.459"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "9 USD"] <- "$6.453"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "9.5 USD"] <- "$6.812"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "23 NZD"] <- "$11.597"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "28 USD"] <- "$20.078"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "19 AUD"] <- "$10.369"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "20 USD"] <- "$14.342"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "$14.600 aprox."] <- "$14.600"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "14.90 soles (aprox $3.000)"] <- "$3.000"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "7 USD"] <- "$5.019"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "16 USD"] <- "$11.473"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "16.9 AUD en pan normal, y 66.9 en pan de oro"] <- "$9.223"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "14.00 NZD"] <- "$7.059"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "$4.200 con bebida"] <- "$4.200"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "13.5 AUD"] <- "$7.368"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == " 26 USD"] <- "$18.644"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "$140 pesos argentinos (Aprox $3.150)"] <- "$3.150"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "13 USD"] <- "$9.322"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "14 AUD"] <- "$7.640"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "25 USD"] <- "$17.927"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "17 USD"] <- "$12.190"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "22 NZD"] <- "$11.093"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "18 AUD"] <- "$9.824"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "7 USD"] <- "$5.019"

Después se elimina el signo de peso ($) para que se pueda leer como integer

ing\_cbind\_precios\_correctos$Precio = as.numeric(gsub("\\$", "", ing\_cbind\_precios\_correctos$Precio))

## Warning: NAs introduced by coercion

Para poder hacer un análisis de mejor calidad, decidí que sería mejor elegir un ingrediente principal y enfocarme solamente en ese. Elegí la hamburguesa de carne/vacuno. Era necesario limpiar los datos del nombre del ingrediente, ya que viene escrito de muchas maneras distintas en el dataframe.

ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == " Hamburguesa"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == " Hamburguesa con queso azul gratinado"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == " Hamburguesa de 180 grs de punta picana y tocino"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == " Hamburguesa de 200 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == " Hamburguesa de carne de vacuno y cerdo molida condimentada con ajo y aj�  
"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "$Hamburguesa"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "3 Hamburguesas de mezcla de wagyu con angus  
"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Burger casera de 120 gramos"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Cheeseburger con pebre (palta"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Doble hamburguesa"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Doble hamburguesa (220 grs)"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hambuguesa"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa (340 grs de carne)"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa 50% carne y 50% prieta"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa a la parrilla"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa al carb�n"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa Angus de 230 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa Angus rellana con queso"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa Black Angus"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa casera"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa casera a la parrilla"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa casera de 150 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa casera de 170 gramos rellena con queso Cheddar"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa con queso Brie apanado"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa con queso mantecoso fundido"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa con tocino"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa cubierta con crema de queso azul"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 150 gramos"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 180 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 180 grs rellena con queso Cheddar"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 200 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 200 grs rellena de queso Cheddar"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 220 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 225 grs."] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 230 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 250 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 3 cortes de carne"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de 300 gramos rellena de queso"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de ara�ita de Wagyu"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de vacuno (posta)"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de vacuno Hereford de 175 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de Wagyu"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de wagyu de 185 grs"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa de Wagyu y Angus con cebolla caramelizada al maple syrup"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa dry-aged"] <- "Hamburguesa de vacuno"  
ing\_cbind\_precios\_correctos[ing\_cbind\_precios\_correctos == "Hamburguesa Hereford"] <- "Hamburguesa de vacuno"

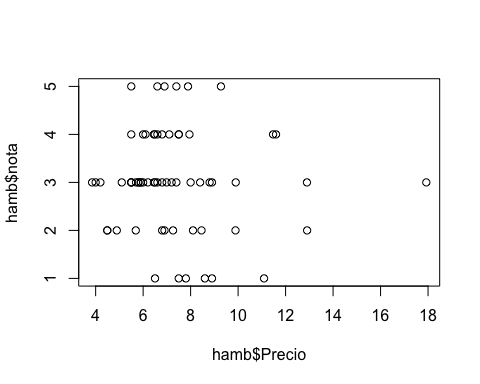
Lo ultimo que falta para limpiar los datos sería mostrar solamente los sandwiches de hamburguesa de vacuno

colnames(ing\_cbind\_precios\_correctos)[colnames(ing\_cbind\_precios\_correctos) == "1"] <- "ingrediente\_principal"  
hamb<- filter(ing\_cbind\_precios\_correctos, ingrediente\_principal == "Hamburguesa de vacuno")

Ahora que tenemos el dataframe limpio y solamente con los datos necesarios podemos comenzar el análisis de los datos para encontrar el sándwich de hamburguesa que asegura una buena nota.

Primero hice un análisis para ver la relación entre los precios y las notas de los sandwiches. Para esto use un scatter plot

relacion<- plot(hamb$Precio, hamb$nota)



relacion

## NULL

Como se puede ver en el scatter plot, el precio es bastante irrelevante en cuanto a la nota que se le da al sandwich.

Al mismo tiempo, al revisar una muestra de 5 sandwiches de nota 1, nota 3 y nota 5, note que los ingredientes varían mucho entre los sandwiches, y no existe una consistencia en la relación con sus notas. Para investigar más, leí los textos de cada crítica, y fue ahí donde empecé a encontrar patrones. Más allá de que ingredientes tiene cada sandwich, me di cuenta que lo importante era la calidad de los ingredientes, especialmente la calidad de la carne y del pan. Con respecto a la carne, era muy importante su punto de cocción y si era jugosa. Con respecto al pan era clave que fuera casero o al menos de muy buena calidad. Aunque los otros “toppings” se repetían mucho entre las notas malas y notas buenas, lo que sí se distingue es su cantidad (o presencia en el sandwich).