Assignment 1: The cocktail bar (data transformation and manipulation)

Tamas Nagy

# Skills needed to solve this assignment

* Using R and RStudio, reading data
* Reporting using RMarkdown
* Using Git and Github (for submitting the task)
* Data manipulation (e.g. dplyr, tidyr)
* String manipulation (e.g. stringr)

# Background

Your Cuban uncle sets up a cocktail bar in downtown Budapest. He has a secret list of cocktails that he would like to serve in the bar. He asks you to do create a few lists and tables to set up the shop. As your uncle is a secret hipster, he has a dataset on Github that you can use for the task!



Note: There are several different ways to solve these tasks, feel free to come up with your own.

## 1. Read the data

Read the cocktail dataset from: <https://github.com/nthun/cocktail-balance> You can also find the codebook there.

cocktails <- read\_tsv("https://raw.github.com/nthun/cocktail-balance/master/cocktail\_data.tsv")

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## name = col\_character(),  
## abv = col\_double(),  
## acid = col\_double(),  
## sugar = col\_double(),  
## type = col\_character(),  
## index = col\_double(),  
## instructions = col\_character(),  
## ingredients = col\_character(),  
## ncotw = col\_character()  
## )

## 2. Transform the data table and clean the ingredient variable!

The ingredients are currently in a single cell for each cocktail. It would be better to put them in separate rows. Also, the variable should be cleaned of all quantities (e.g. 1/2 oz, 2 dashes, etc.), and the indicator of alcohol content (e.g. 47.3% abv). You will need to use this cleaned table in later tasks.

cocktails\_t <- cocktails %>%  
 separate\_rows("ingredients", sep = "<br>")%>%  
 separate\_rows("ingredients", sep = "<br/>")  
cocktails\_clean <- cocktails\_t %>%  
 mutate(ingredients\_clean = str\_remove\_all(ingredients, "[:digit:]") %>%  
 str\_remove\_all("oz") %>%  
 str\_remove\_all("drops") %>%  
 str\_remove\_all("dash") %>%  
 str\_remove\_all("%") %>%  
 str\_remove\_all("(. abv)") %>%  
 str\_remove\_all("[(]")%>%  
 str\_remove\_all("[)]")%>%  
 str\_remove\_all("[/]") %>%  
 str\_remove\_all("[.]")%>%  
 str\_squish())

## 3. All ingredients in alphabetical order

Before opening the bar, you need to find a reliable supplier that has all the ingredients. You need to send a list of all possible ingredients you will need. They don’t need the quantities (i.e. how many of these are needed), just the names of the ingredients.

ingredients\_list <- cocktails\_clean %>%  
 arrange(ingredients\_clean)  
unique(ingredients\_list$ingredients\_clean)

## [1] "Absolut Citron vodka"   
## [2] "agar-clarified grapefruit juice"   
## [3] "Amer Picon"   
## [4] "Angostura bitters"   
## [5] "apple brandy"   
## [6] "Benedictine"   
## [7] "blanco tequila"   
## [8] "bourbon"   
## [9] "bsp Fernet Branca"   
## [10] "bsp raspberry syrup"   
## [11] "Campari"   
## [12] "centrifuge-clarified grapefruit juice"   
## [13] "Cherry Herring"   
## [14] "clarified lime juice"   
## [15] "clarified lime juice or champagne acid"  
## [16] "Cognac"   
## [17] "Cointreau"   
## [18] "cranberry juice"   
## [19] "crem de violette"   
## [20] "Curacao"   
## [21] "Demerara syrup"   
## [22] "Dolin dry vermouth"   
## [23] "Drambuie"   
## [24] "drop saline"   
## [25] "dry vermouth"   
## [26] "egg white"   
## [27] "es absinthe"   
## [28] "es Angostura bitters"   
## [29] "es champagne acid"   
## [30] "es orange bitters"   
## [31] "es Peychaud's bitters"   
## [32] "gin"   
## [33] "grapefruit juice"   
## [34] "Green Chartreuse"   
## [35] "Grenadine"   
## [36] "heavy cream"   
## [37] "Hellfire bitters"   
## [38] "honey syrup"   
## [39] "Lairds Applejack Bottled in Bond"   
## [40] "lemon juice"   
## [41] "Lillet Blanc"   
## [42] "lime juice"   
## [43] "Luxardo Maraschino"   
## [44] "Luxardo Marschino<b> Plymouth gin"   
## [45] "Old Tom gin"   
## [46] "orange bitters"   
## [47] "orange juice"   
## [48] "Peychaud's bitters"   
## [49] "pisco"   
## [50] "Plymouth gin"   
## [51] "quinine simple syrup"   
## [52] "raspberry syrup"   
## [53] "rye"   
## [54] "saline solution"   
## [55] "Scotch"   
## [56] "simple"   
## [57] "simple syrup"   
## [58] "sloe gin"   
## [59] "sugard -proof rum"   
## [60] "sugared -proof rye"   
## [61] "sweet vermouth"   
## [62] "water"   
## [63] "white cr<U+008F>me de cacao"   
## [64] "white mezcal"   
## [65] "white rum"   
## [66] "Yellow Chartreuse"

## 4. Number of unique ingredients

How many different ingredients you will need?

n\_distinct(cocktails\_clean$ingredients\_clean)

## [1] 66

## 5. What are the top 10 ingredients?

What are the 10 most frequently used ingredients? If there are ties, you can list more than 10.

sort(table(cocktails\_clean$ingredients\_clean),decreasing=TRUE)[1:26]

##   
## lemon juice simple syrup gin   
## 15 14 13   
## sweet vermouth lime juice saline solution   
## 13 12 10   
## water rye Angostura bitters   
## 10 8 6   
## Cognac es Angostura bitters Luxardo Maraschino   
## 6 6 6   
## bourbon Cointreau egg white   
## 5 5 5   
## white rum Benedictine clarified lime juice   
## 5 4 4   
## Green Chartreuse honey syrup Plymouth gin   
## 4 4 4   
## Scotch Campari Curacao   
## 4 3 3   
## grapefruit juice orange bitters   
## 3 3

## 6. Which cocktail(s) has/have the most ingredients?

Count the number of ingredients and filter all the cocktails that has that many.

sort(table(cocktails\_clean$name), decreasing=T)

##   
## Blender Margarita Carbonated Negroni   
## 6 6   
## Clover Club De La Louisiane   
## 6 6   
## Pink Lady Vieux Carre   
## 6 6   
## Blender Whiskey Sour Brooklyn   
## 5 5   
## Carbonated Margarita Carbonated Whiskey Sour   
## 5 5   
## Champs-Elysses Gin and Juice Centrifuge   
## 5 5   
## Gin and Tonic (Dry) Hemingway Daiquiri   
## 5 5   
## Margarita Martinez   
## 5 5   
## Pegu Club Whiskey Sour   
## 5 5   
## 20th Century Cocktail Aviation   
## 4 4   
## Between the Sheets Bijou   
## 4 4   
## Blackthorn Blender Daiquiri   
## 4 4   
## Blood and Sand Brady Crusta   
## 4 4   
## Corpse Reviver #2 Cosmopolitan (Modern/Bastardized)   
## 4 4   
## Gin and Juice Agar Improved Whiskey Cocktail   
## 4 4   
## Jack Rose Last Word   
## 4 4   
## Pisco Sour Sidecar   
## 4 4   
## Alexander Bee's Knees   
## 3 3   
## Blinker Bobby Burns   
## 3 3   
## Brown Derby Chartruth   
## 3 3   
## Classic Daiquiri Daiquiri with More Lime   
## 3 3   
## Fresh Lime Gimlet Gold Rush   
## 3 3   
## Hanky Panky Honeysuckle   
## 3 3   
## Manhatan (Boubon, 45% abv) Manhatan (Rye, 50% abv)   
## 3 3   
## Negroni Old Fashioned   
## 3 3   
## Old Pal Rob Roy   
## 3 3   
## Southside Widow's Kiss   
## 3 3   
## Rusty Nail   
## 2

cocktails\_n <- cocktails\_clean %>%  
 group\_by(name) %>%  
 mutate(freq=n())%>%  
 arrange(desc(freq))  
  
unique(cocktails\_n$name)

## [1] "Pink Lady" "Clover Club"   
## [3] "Vieux Carre" "De La Louisiane"   
## [5] "Blender Margarita" "Carbonated Negroni"   
## [7] "Whiskey Sour" "Hemingway Daiquiri"   
## [9] "Margarita" "Champs-Elysses"   
## [11] "Pegu Club" "Martinez"   
## [13] "Brooklyn" "Blender Whiskey Sour"   
## [15] "Carbonated Margarita" "Carbonated Whiskey Sour"   
## [17] "Gin and Tonic (Dry)" "Gin and Juice Centrifuge"   
## [19] "Pisco Sour" "Blood and Sand"   
## [21] "Cosmopolitan (Modern/Bastardized)" "20th Century Cocktail"   
## [23] "Corpse Reviver #2" "Jack Rose"   
## [25] "Aviation" "Sidecar"   
## [27] "Last Word" "Between the Sheets"   
## [29] "Brady Crusta" "Blackthorn"   
## [31] "Bijou" "Improved Whiskey Cocktail"   
## [33] "Blender Daiquiri" "Gin and Juice Agar"   
## [35] "Daiquiri with More Lime" "Honeysuckle"   
## [37] "Classic Daiquiri" "Alexander"   
## [39] "Brown Derby" "Gold Rush"   
## [41] "Southside" "Bee's Knees"   
## [43] "Fresh Lime Gimlet" "Blinker"   
## [45] "Negroni" "Hanky Panky"   
## [47] "Manhatan (Boubon, 45% abv)" "Bobby Burns"   
## [49] "Rob Roy" "Old Pal"   
## [51] "Manhatan (Rye, 50% abv)" "Widow's Kiss"   
## [53] "Old Fashioned" "Chartruth"   
## [55] "Rusty Nail"

max\_ingredients <- cocktails\_n %>%  
 filter(freq == max(cocktails\_n$freq))  
  
unique(max\_ingredients$name)

## [1] "Pink Lady" "Clover Club" "Vieux Carre"   
## [4] "De La Louisiane" "Blender Margarita" "Carbonated Negroni"

## 7. How many ingredients appear in only one cocktail (rare ingredient)?

Count all cocktails with only one ingredient, and

sort(table(cocktails\_clean$ingredients\_clean), decreasing = F)

##   
## Absolut Citron vodka agar-clarified grapefruit juice   
## 1 1   
## Amer Picon bsp Fernet Branca   
## 1 1   
## bsp raspberry syrup centrifuge-clarified grapefruit juice   
## 1 1   
## Cherry Herring clarified lime juice or champagne acid   
## 1 1   
## cranberry juice crem de violette   
## 1 1   
## Demerara syrup Dolin dry vermouth   
## 1 1   
## Drambuie drop saline   
## 1 1   
## es absinthe es champagne acid   
## 1 1   
## es orange bitters es Peychaud's bitters   
## 1 1   
## heavy cream Hellfire bitters   
## 1 1   
## Lairds Applejack Bottled in Bond Luxardo Marschino<b> Plymouth gin   
## 1 1   
## Old Tom gin Peychaud's bitters   
## 1 1   
## pisco quinine simple syrup   
## 1 1   
## raspberry syrup simple   
## 1 1   
## sloe gin sugard -proof rum   
## 1 1   
## sugared -proof rye white cr<U+008F>me de cacao   
## 1 1   
## white mezcal apple brandy   
## 1 2   
## blanco tequila dry vermouth   
## 2 2   
## Grenadine Lillet Blanc   
## 2 2   
## orange juice Yellow Chartreuse   
## 2 2   
## Campari Curacao   
## 3 3   
## grapefruit juice orange bitters   
## 3 3   
## Benedictine clarified lime juice   
## 4 4   
## Green Chartreuse honey syrup   
## 4 4   
## Plymouth gin Scotch   
## 4 4   
## bourbon Cointreau   
## 5 5   
## egg white white rum   
## 5 5   
## Angostura bitters Cognac   
## 6 6   
## es Angostura bitters Luxardo Maraschino   
## 6 6   
## rye saline solution   
## 8 10   
## water lime juice   
## 10 12   
## gin sweet vermouth   
## 13 13   
## simple syrup lemon juice   
## 14 15

ingredients\_n <- cocktails\_clean %>%  
 group\_by(ingredients\_clean) %>%  
 mutate(freq=n())%>%  
 arrange(desc(freq))  
  
unique\_ingredients <- ingredients\_n %>%  
 filter(freq == min(ingredients\_n$freq))  
n\_distinct(unique\_ingredients$name)

## [1] 24

## 8. Which cocktail has an ingredient that is only used in one cocktail?

unique(unique\_ingredients$name)

## [1] "Pisco Sour" "Pink Lady"   
## [3] "Clover Club" "Blood and Sand"   
## [5] "Alexander" "Cosmopolitan (Modern/Bastardized)"  
## [7] "20th Century Cocktail" "Aviation"   
## [9] "Last Word" "Blinker"   
## [11] "Blackthorn" "Hanky Panky"   
## [13] "Martinez" "Vieux Carre"   
## [15] "Brooklyn" "Rusty Nail"   
## [17] "De La Louisiane" "Blender Daiquiri"   
## [19] "Blender Margarita" "Blender Whiskey Sour"   
## [21] "Gin and Tonic (Dry)" "Carbonated Negroni"   
## [23] "Gin and Juice Centrifuge" "Gin and Juice Agar"

## 9. What are the cocktails without rare ingredients?

nonunique\_ingredients <- ingredients\_n %>%  
 filter(freq > 1)  
unique(nonunique\_ingredients $name)

## [1] "Pink Lady" "Clover Club"   
## [3] "Whiskey Sour" "Gold Rush"   
## [5] "Southside" "Bee's Knees"   
## [7] "20th Century Cocktail" "Corpse Reviver #2"   
## [9] "Jack Rose" "Aviation"   
## [11] "Sidecar" "Champs-Elysses"   
## [13] "Between the Sheets" "Brady Crusta"   
## [15] "Blender Whiskey Sour" "Pisco Sour"   
## [17] "Daiquiri with More Lime" "Classic Daiquiri"   
## [19] "Fresh Lime Gimlet" "Margarita"   
## [21] "Improved Whiskey Cocktail" "Old Fashioned"   
## [23] "Carbonated Margarita" "Carbonated Whiskey Sour"   
## [25] "Blood and Sand" "Pegu Club"   
## [27] "Negroni" "Blackthorn"   
## [29] "Hanky Panky" "Martinez"   
## [31] "Manhatan (Boubon, 45% abv)" "Bobby Burns"   
## [33] "Rob Roy" "Vieux Carre"   
## [35] "Bijou" "Manhatan (Rye, 50% abv)"   
## [37] "De La Louisiane" "Gin and Tonic (Dry)"   
## [39] "Carbonated Negroni" "Gin and Juice Centrifuge"   
## [41] "Gin and Juice Agar" "Honeysuckle"   
## [43] "Hemingway Daiquiri" "Cosmopolitan (Modern/Bastardized)"  
## [45] "Last Word" "Blender Daiquiri"   
## [47] "Blender Margarita" "Chartruth"   
## [49] "Blinker" "Old Pal"   
## [51] "Brooklyn" "Alexander"   
## [53] "Brown Derby" "Rusty Nail"   
## [55] "Widow's Kiss"

## 10. Create a cheat sheet for the bartender!

Create a matrix that shows all cocktail names as rows and all ingredients as columns. When a cocktail requires an ingredient, there should be an “X” in the cell, otherwise, the cell should remain empty. Example:

Congrats, the bar is now officially open!

