

# Fried food consumption in demographic groups

Syona Neeraj

26/05/2021

## Data Origins

The data was collected from online open sources and can be viewed here (<https://www.kaggle.com/jleibow27/fried-food-consumption-and-mortality>). The full dataset records the consumption of fried food in different demographic groups based on the number of servings consumed during a week.

## Research Question

Fried food consumption was tracked in women from the US, belonging to different demographic groups. Here, we analyze the trend of number of *fried food servings* consumed along the week within different groups.

This could point out the cultural norms and habits as followed within them. Comparing them to the trend observed in baseline disorders would provide insight into how the consumption affects different groups.

However, it is important to note that the baseline disorders observed are not correlated to the trend in racial groups so we cannot conclude how the consumption in different groups would directly lead to disorders.

1. Set working directory and load all the necessary libraries
2. Once the libraries are ready, we load the data and view it for clarity.

```
Data <- read.csv(here("Data", "Fried food.csv")) #Loaded the data
View(Data)
head(Data)
```

3. We clean the data in order to use it efficiently. The P values are not necessary, neither is the age at baseline as they are not relevant to our objective.

```
Data <- Data[-c(1),]
Data <- Data%>% select(-P)
Data <- Data[-c(1),]
names(Data)
```

```
## [1] "Factors" "None..n.15.166."
## [3] "X1.1.serving.week..n.38.482." "X1.2.servings.week..n.33.210."
## [5] "X3.6.servings.week..n.15.480." "â..1.serving.day..n.4628."
```

4. For convenience, we rename the columns to more accessible names as follows :

where, **serving\_0 = 0 servings per week**

**serving\_1 = 0 to 1 servings per week**

**serving\_2 = 1 to 2 servings per week**

**serving\_3 = 3 to 6 servings per week**

## serving\_max = more than 1 serving per day

```
names(Data2)
```

```
## [1] "Factors"      "serving_0"    "serving_1"    "serving_2"    "serving_3"
## [6] "serving_max"
```

5. Next, we must create subsets of the data that we do require. There are two groups that we need:

A. Fried food in racial groups (Rows 1 to 5)

B. Disorders observed (Rows 39 to 41)

```
dfrace <- Data2 %>% slice_head(n = 5)
df_diseases <- Data2 %>% slice(39:41)
```

```
as.matrix.data.frame(dfrace)
```

```
##      Factors    serving_0 serving_1 serving_2 serving_3 serving_max
## [1,] "White"    "90.5"      "87.9"    "82.2"    "77.1"    "68.5"
## [2,] "Black"    "4.4"       "5.5"     "8.5"     "11.3"    "17.2"
## [3,] "Hispanic" "2.4"       "3"       "4.2"     "5.7"     "7.7"
## [4,] "Others"   "2.5"       "3.3"     "5"       "5.8"     "6.5"
## [5,] "Missing" "0.3"       "0.3"     "0.2"     "0.2"     "0.2"
```

```
dfrace[, -1]
```

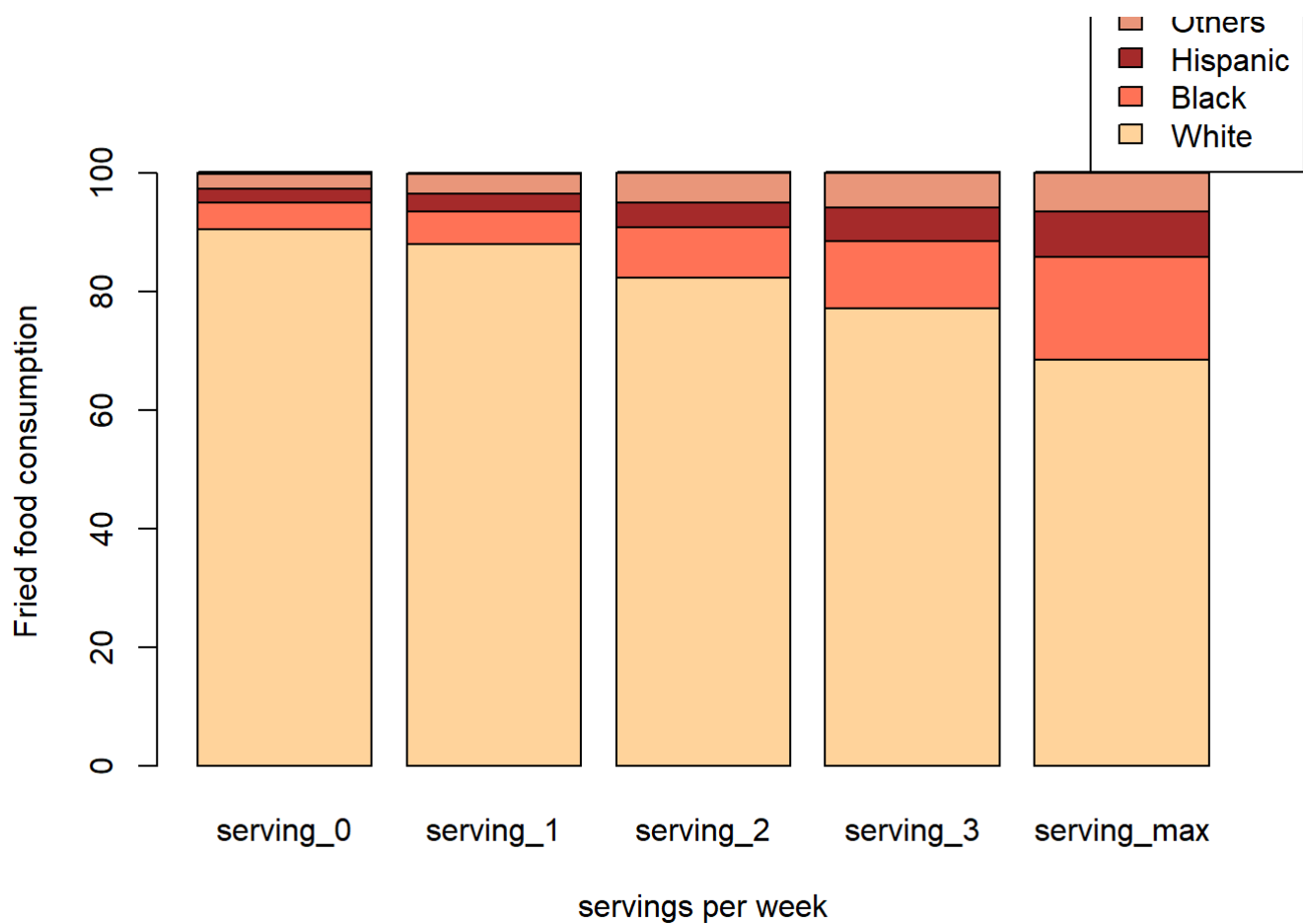
<b>serving_0</b> <chr>	<b>serving_1</b> <chr>	<b>serving_2</b> <chr>	<b>serving_3</b> <chr>	<b>serving_max</b> <chr>
90.5	87.9	82.2	77.1	68.5
4.4	5.5	8.5	11.3	17.2
2.4	3	4.2	5.7	7.7
2.5	3.3	5	5.8	6.5
0.3	0.3	0.2	0.2	0.2

5 rows

```
mytable1 = as.matrix(dfrace[, -1])
```

## Visualization 01

```
barplot(mytable1,
  xlab="servings per week",
  ylab= "Fried food consumption",
  legend.text = dfrace$Factors,
  col = c("burlywood1","coral1","brown","darksalmon","gold"),
  args.legend = list(x = "bottomright",inset = c(- 0.05, 1)))
```



We follow the same routine for the graph plotting the disorders:

```
as.matrix.data.frame(df_diseases)
```

```
##      Factors                                serving_0 serving_1 serving_2 serving_3
## [1,] "Baseline diabetes"                    "4.5"      "4.9"      "5.8"      "7.3"
## [2,] "Baseline cardiovascular disease"       "7"        "6.3"      "6.3"      "5.8"
## [3,] "Baseline cancer"                      "1.5"      "1.5"      "1.5"      "1.6"
##      serving_max
## [1,] "9.7"
## [2,] "6.1"
## [3,] "1.3"
```

```
df_diseases[, -1]
```

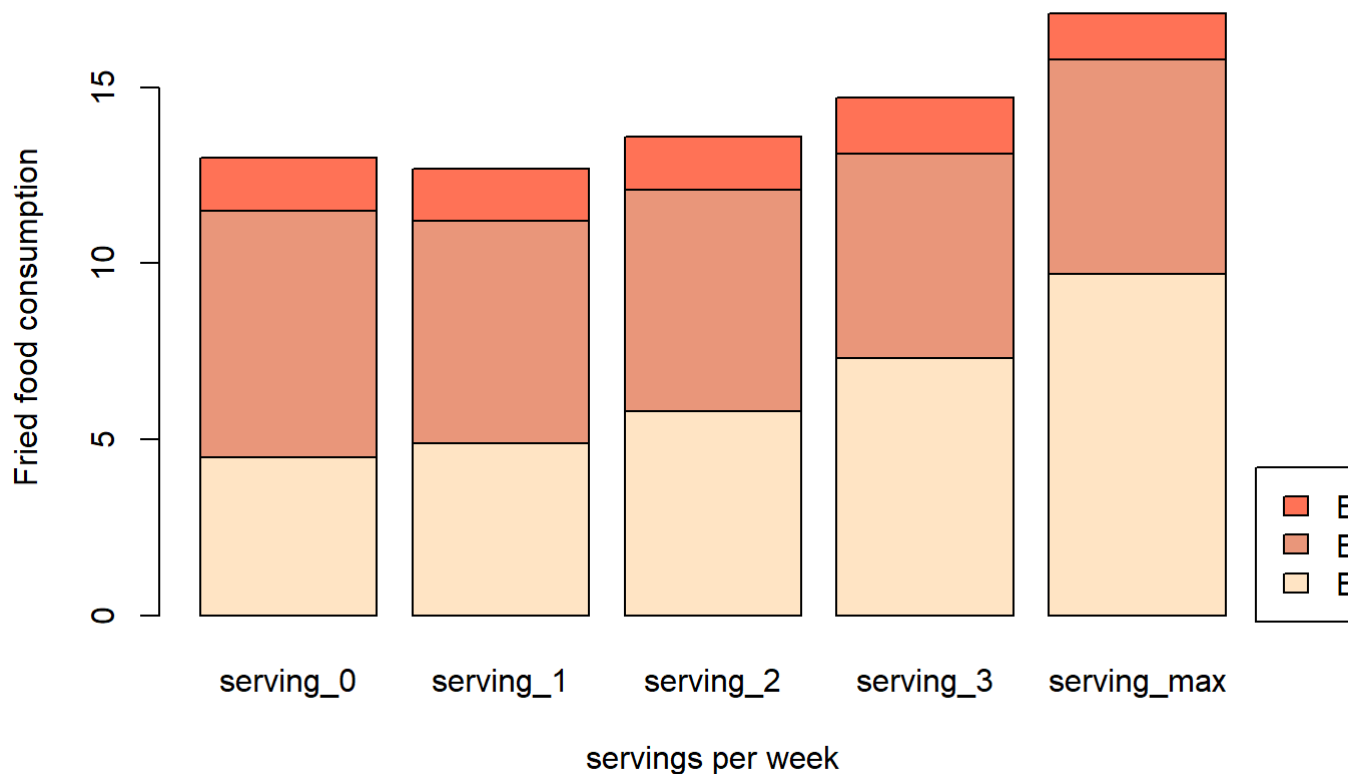
serving_0 <chr>	serving_1 <chr>	serving_2 <chr>	serving_3 <chr>	serving_max <chr>
4.5	4.9	5.8	7.3	9.7
7	6.3	6.3	5.8	6.1
1.5	1.5	1.5	1.6	1.3

3 rows

```
mytable2 = as.matrix(df_diseases[, -1])
```

## Visualization 02

```
barplot(mytable2,  
  xlab = "servings per week",  
  ylab = "Fried food consumption",  
  legend.text=df_diseases$Factors,  
  col=c("bisque1","darksalmon","coral1"),  
  args.legend = list(x = "bottomright",inset = c(- 0.48, 0)))
```



## Summary

Through the visualization performed above, we can conclude that White women consume the highest level of fried food, followed by Black and Hispanic women. The rest of the data present accounts for others and missing data

We can also conclude that baseline diabetes is most common among women consuming fried food followed by cardiovascular disorders and finally cancer.

## Caveats

- A. The dataset was present in wider format and was hard to process considering it was spread along columns
- B. The values were present in percentages and made it extremely hard to manipulate the data as the number of participants for each group was different and a standard formula using the mutate function was hard to achieve.

## References

Healy, K. (2018). ***Data visualisation: a practical introduction***. (<https://socviz.co/>)Princeton University Press.

*Association of fried food consumption with all cause, cardiovascular, and cancer mortality*  
(<https://www.bmj.com/content/364/bmj.k5420>)