

DATA 2010 Group Project

Lucas Berzuk, Ethan Robson, Hugo Ng, Rodell Salonga

Due on 015/03/2024

This data comes from a study performed in Mexico, Peru and Colombia. It has 17 attributes and 2111 data points. It is the study of how peoples eating habits and their physical condition has an effect on their level of obesity. It is important to note that up to 77% of this data has been synthetically generated because of a balancing issue so this is not all 100% real data. There are many variables that have unclear descriptions as headers so here is a list for clarity. Frequent consumption of high caloric food (FAVC), Frequency of consumption of vegetables (FCVC), Number of main meals (NCP), Consumption of food between meals (CAEC), Consumption of water daily (CH20), and Consumption of alcohol (CALC). The attributes related with the physical condition are: Calories consumption monitoring (SCC), Physical activity frequency (FAF), Time using technology devices (TUE), Transportation used (MTRANS)

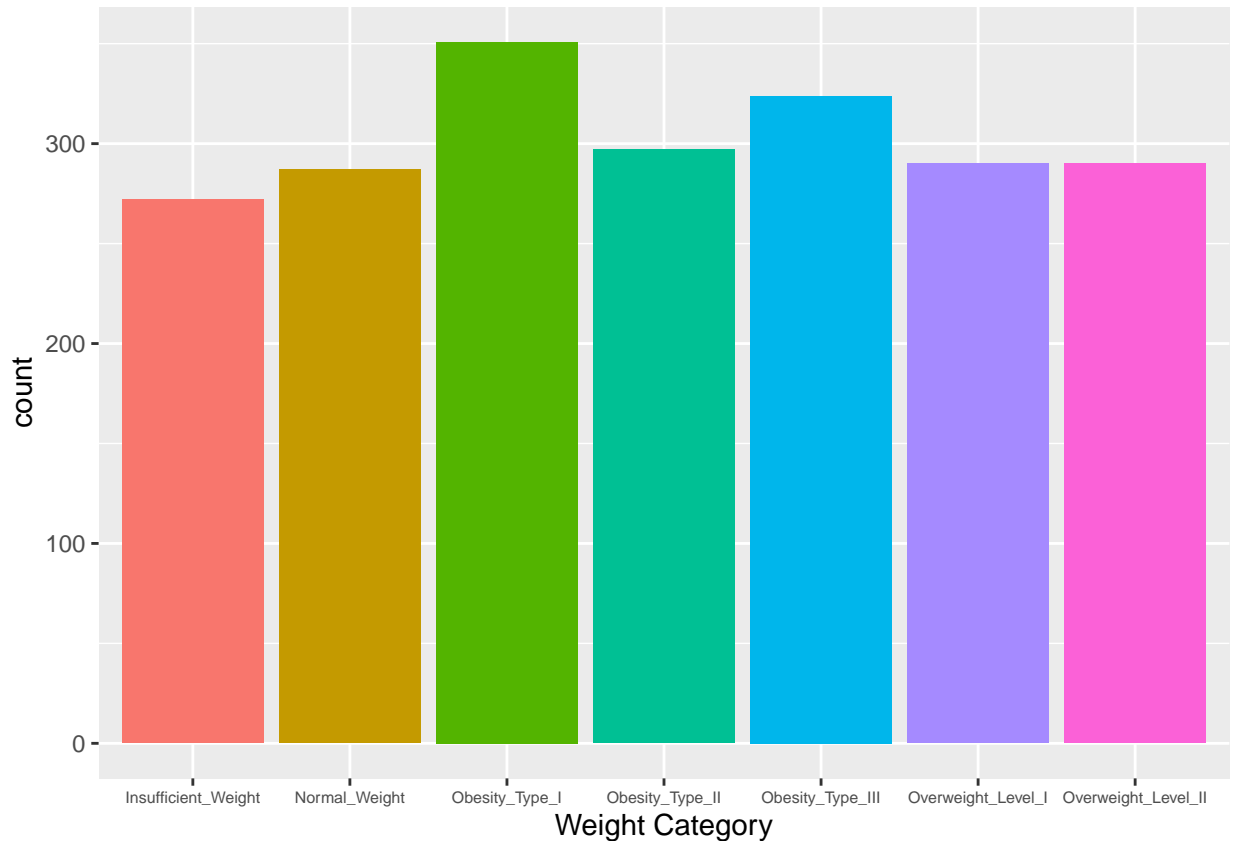
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
ObesityDataset = read.csv("ObesityDataSet_raw_and_data_sinthetic.csv")
```

```
# A Plot of how many people analyzed were in each weight category
```

```
ObesityDataset |> ggplot(aes(x = NObeyesdad, fill = NObeyesdad)) + geom_bar() +  
  theme(axis.text.x=element_text(size=6), legend.position = "none") + xlab("Weight Category")
```



Where we currently are with analyzing the data is seeing how many people fall under each weight category and what the reasoning is. Each category is relatively even based on the plot above so we know now that there are many data points in each weight category that we can analyze to figure out why people are overweight (or at least what factors correlate to being overweight). What we have already found out (based on another plot that is not included here) is that the amount of people that ‘sometimes’ drink alcohol are always the majority in every weight category, with the people that answered ‘no’ always being the second highest answer. Minimal people have answered ‘frequently’ and ‘always’ but this data maybe be skewed because we don’t expect people that have a drinking problem wanting to own up to their drinking problem. We also noticed that in the ‘Overweight_Level_I’, ‘Obesity_Type_II’ and ‘Obesity_Type_III’ Category, the number of people that answer ‘sometimes’ for alcohol is significantly higher than the number of people that answer ‘no’, compared to the rest of the categories.

The direction that we want to go with this is... (we need to come to a conclusion here