Group meeting time:
Sunday 1-2 pm

Mode of Collaboration:
We have a group chat where we can share all the files including PDFs, RMDs, as well as CSV files. We will divide project work assignments so that every team member essentially has an equal proportion of the workload. During our meeting time, we will go over logistics first and discuss the progress of our project.

First Dataset:

Name of dataset:
[transit_cost.csv]
Source of dataset:
[We found the data on GitHub provided by professor.]
Link to dataset:
[https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-01-05/ readme.md]
A short paragraph (3-6 sentences) explaining why you picked this dataset as a potential candidate for the final project:
[We chose this dataset because it satisfies the requirements as it contains both categorical variables like city, line, start_year, country, and quantitative variables like length, tunnel, stations, tunnel_per, coast, etc. We found this dataset meaningful because by exploring the underlying correlations behind the dataset, we might help to reduce the cost per kilometer basis. The budget we saved can be used in more needy areas like education. We are curious about whether the amount of years used in constructing the tunnel is correlated to the difference in cost per km in millions. Since by intuition the low efficiency of workers might be part of the reason of extremely high budget cost.]

Second Dataset:
Name of dataset:
[Methane Emissions Around The World (1990-2018)]
Source of dataset:
[We found the data on Kaggle provided by Climate Watch Data]
Link to dataset:
[https://www.kaggle.com/datasets/kkhandekar/methane-emissions-across-the-world-19902018]
A short paragraph (3-6 sentences) explaining why you picked this dataset as a potential candidate for the final project:
[We chose this dataset because it satisfies the requirements as it contains both categorical variables, such as country and sector, and continuous variables, like methane emission level each year from 2013 to 1990. There are 195 countries in this dataset. With this many variables, there's

a variety of analyses that can be done in this dataset. From this dataset, we can determine which country has the most methane emission, and what factors would cost the most methane emission.]