

Homework 3 Part 1 Appendix

Abdulwahab, Fatma

Barrett, Alexis

McDonough, Jared

Miller, Matthew

Solares, Nicky

March 4, 2019

1 Journey Length

The total distance was 6425 miles.

2 Trip Map

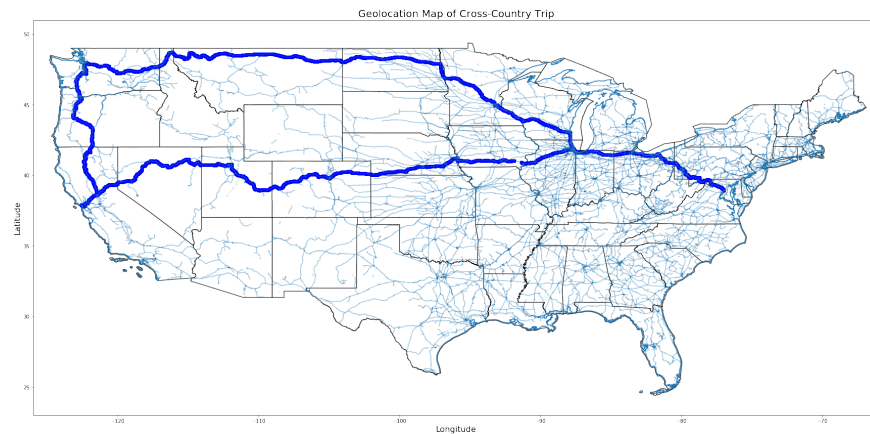


Figure 1: Trip Map

3 Visited Cities

4 Total Stops and the Definition of a Stop

5 Plotted Stops

6 Data Between Stops

The average speed between stops is 50.82 mph, with a standard deviation of 24.14. In kilometers, it is 81.79kmh with a standard deviation of 38.85.

7 Visited States

The states visited include Virginia, West Virginia, Pennsylvania, Ohio, Maryland, California, Nevada, Indiana, Illinois, Wisconsin, Minnesota, North Dakota, Montana, Idaho, Washington, Oregon, Utah, Colorado, Kansas, and Nebraska

8 Landmarks

9 Appendix Code: Distance

```
import pandas as pd
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
import scipy as sp
import geopandas
import geoplot
import geoplot.crs as gcrs
from shapely.geometry import Point, Polygon
import math
'''

from sklearn import linear_model
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import LabelEncoder, OneHotEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
'''

plt.style.use('dark_background')

#Tidy the timestamp data
time_list = []
for i in CC_Data["timestamp"]:
    '''This loop separates the clock time from the rest of the string'''
    start = 0 #index for slicing
    stop = 0 #index for slicing
    flag = 0
    for j in range(len(i)):
        if (i[j] == " " and flag == 0):
            start = j+1
            flag = 1
        elif (i[j] == " " and flag == 1):
            stop = j
    time_list.append(i[start:stop])

CC_Data["time"] = time_list

second_list = []
for time in CC_Data["time"]:
    '''this will create a column of all the times in seconds....this is easier to work with'''
```

```

x = time.split(":")
hour = int(x[0])
minute = int(x[1])
second = int(x[2])
second_list.append((hour%12)*60^2 + minute*60 + second)

for i in range(len(second_list)):
'''This changes the list so that each time is relative to the start time'''
second_list[i] -= 574

CC_Data["seconds"] = second_list

#remove points where the accuracy >10
CC_Data = CC_Data[CC_Data["accuracy"] <= 10]

def Haversine(lat1, lon1, lat2, lon2):
R = 6372.8 # Earth radius in kilometers

dLat = np.radians(lat2 - lat1)
dLon = np.radians(lon2 - lon1)

lat1 = np.radians(lat1)
lat2 = np.radians(lat2)

a = np.sin(dLat/2)**2 + np.cos(lat1)*np.cos(lat2)*np.sin(dLon/2)**2
c = 2 * np.arctan2(np.sqrt(a), np.sqrt(1-a))

return R * c

Total_Dist = 0.0
for i in range(len(CC_Data)-1):
Total_Dist += Haversine(lat1=CC_Data["latitude"].iloc[i], lon1=CC_Data["longitude"].iloc[i],
lat2=CC_Data["latitude"].iloc[i+1], lon2=CC_Data["longitude"].iloc[i+1])

print("The total distance covered in this Cross Country was %0.3f Km or %0.3f miles" %(Total_Dist, Total_Dist*0.621371))

```

10 Appendix Code: Between Stop Data

```

import pandas as pd

df = pd.read_csv("dataset1_combined.csv")
dfclean = df[df.accuracy <= 20]

nonstopmphdf = df[(df['speedmph'] != 0)]
print(nonstopmphdf['speedmph'].mean())
print(nonstopmphdf['speedmph'].std())

nonstopkmhdf = df[(df['speedkmh'] != 0)]
print(nonstopkmhdf['speedkmh'].mean())
print(nonstopkmhdf['speedkmh'].std())

```