

# Activity 1 - Rivers of the World

Jett Rugebregt

4/4/2024

```
In [1]: import pandas as pd  
import numpy as np
```

```
In [2]: # read in the csv with pandas  
file = pd.read_csv("C:/Users/jettr/Dropbox (University of Oregon)/23-24/Spring/Geog  
  
# Did it read in the file right? What does it look like?  
print(file)  
type(file)
```

	River	Continent	Average discharge (m3/s)	Length (km)	\
0	Amazon	South America	224000	6992	
1	Congo	Africa	41400	4370	
2	Nile	Africa	2810	6853	
3	Mississippi	North America	21300	3766	
4	Rio de la Plata	South America	27225	290	
5	Ob	Asia	13100	4345	
6	Parana	South America	19706	4880	
7	Yenisei	Asia	20200	4090	
8	Lena	Asia	18300	4294	
9	Niger	Africa	7900	4200	
10	Ganges	Asia	43950	3969	
11	Amur	Asia	11526	4444	
12	Yangtze	Asia	31900	6300	
13	Mackenzie	North America	9800	4241	
14	Irtysk	Asia	2980	4248	
15	Volga	Europe	8220	3531	
16	Madeira	South America	30173	3380	
17	Missouri	North America	2445	3767	
18	Saint Lawrence	North America	16800	500	
19	Zambezi	Africa	4217	2574	
20	Ganges	Asia	18691	2525	
21	Paraguay	South America	4696	2621	

	Drainage area (km2)	Outflow	Type
0	6915000	Atlantic Ocean	Primary river
1	3822000	Atlantic Ocean	Primary river
2	3349000	Mediterranean Sea	Primary river
3	3248000	Gulf of Mexico	Primary river
4	3182064	Atlantic Ocean	Primary river
5	2990000	Gulf of Ob	Primary river
6	2582672	Rio de la Plata	Tributary river
7	2580000	Kara Sea	Primary river
8	2490000	Laptev Sea	Primary river
9	2117000	Gulf of Guinea	Primary river
10	1999000	Bay of Bengal	Primary river
11	1855000	Sea of Okhotsk	Primary river
12	1840000	East China Sea	Primary river
13	1805200	Beaufort Sea	Primary river
14	1673470	Ob	Tributary river
15	1380000	Caspian Sea	Primary river
16	1376000	Amazon	Tributary river
17	1371010	Mississippi	Tributary river
18	1344200	Gulf of Saint Lawrence	Primary river
19	1331000	Indian Ocean	Primary river
20	1300000	Bay of Bengal	Primary river
21	1120154	Parana	Tributary river

Out[2]: pandas.core.frame.DataFrame

## Question 1

Number of columns

Number of rows

The maximum Average discharge (m3/s) value

The minimum Drainage area (km2) value

The mean Length (km) value

## a. Number of columns

```
In [3]: # Number of Columns
NumCols = file.shape[1]
print('The Number of Columns : ' + str(NumCols))
```

The Number of Columns : 7

## b. Number of rows

```
In [4]: NumRows = file.shape[0]
print('The Number of Rows : ' + str(NumRows))
```

The Number of Rows : 22

## c. The maximum Average discharge (m3/s) value

```
In [5]: max_discharge = file['Average discharge (m3/s)'].max()
print("Maximum Average Discharge : " + str(max_discharge) + ' m^3 / s')
```

Maximum Average Discharge :224000 m<sup>3</sup> / s

## d. The minimum Drainage area (km2) value

```
In [6]: min_drainage = file['Drainage area (km2)'].min()
print("Minimum drainage area : " + str(min_drainage) + ' km^2')
```

Minimum drainage area : 1120154 km<sup>2</sup>

## e. The mean Length (km) value

```
In [7]: Mean_Length = file['Length (km)'].mean()
print("Average river length : " + str(Mean_Length) + ' km')
```

Average river length : 3917.2727272727275 km

## Question 2

What is the name of the shortest river?

How many of these rivers are located in North America?

What is the mean and standard deviation of Average discharge (m<sup>3</sup>/s)?

What are the names of the rivers that flow into the Atlantic Ocean?

Which continent contains the most large rivers?

## a. What is the name of the shortest river?

```
In [8]: # Find the index where the value of 'Length (km)' is the minimum
shortest_river_index = file['Length (km)'].idxmin()

# What's the name of the River at the index where the minimum Length is
shortest_river_name = file.loc[shortest_river_index, 'River']

# Finding the Length of the Shortest river
Len_Short_River = file.loc[shortest_river_index, 'Length (km)']
print("Name of the shortest river : "+ str(shortest_river_name) + ' with a lenght o
```

Name of the shortest river : Rio de la Plata with a lenght of 290 (km)

## b. How many of these rivers are located in North America?

```
In [9]: # Make a dataframe that is the Continent Column where the cell is 'North America'
NA_rivers = file[file['Continent'] == 'North America']

# How Long is the array we created?
num_NA_rivers = len(NA_rivers)

print("Number of rivers in North America:", num_NA_rivers)
```

Number of rivers in North America: 4

## c. What is the mean and standard deviation of Average discharge (m<sup>3</sup>/s)?

```
In [10]: # Finding the mean and the standard deviation using the pandas package
Mean_AvgDis = file['Average discharge (m3/s)'].mean()
std_AvgDis = file['Average discharge (m3/s)'].std()
print("Mean River Discharge from All Rivers (m^3/s): " + str(Mean_AvgDis) + " and t
```

Mean River Discharge from All Rivers (m<sup>3</sup>/s): 26424.5 and the Standard Deviation (m<sup>3</sup>/s): 45741.92495559033

## d. What are the names of the rivers that flow into the Atlantic Ocean?

```
In [11]: # Make a dataframe where the values within the Outflow Column is 'Atlantic Ocean'
Atlantic_Out_rivers = file[file['Outflow'] == 'Atlantic Ocean']

# Extract the names of rivers flowing into the Atlantic Ocean
Name_At_Out_rivers = Atlantic_Out_rivers['River']

# Convert the pandas Series to a comma-separated string
river_names_str = ', '.join(Name_At_Out_rivers)

print("Rivers flowing out to the Atlantic Ocean :", river_names_str)
```

Rivers flowing out to the Atlantic Ocean : Amazon, Congo, Rio de la Plata

## e. Which continent contains the most large rivers?

```
In [12]: # Count the number of occurrences of each country
country_counts = file['Continent'].value_counts()

# Print the List in order from highest to lowest count
print("Continents from Most to Least rivers :")
print(country_counts)
```

Continents from Most to Least rivers :

Continent

Asia 8

South America 5

Africa 4

North America 4

Europe 1

Name: count, dtype: int64

## Question 3

Add a column called Primary that has value of 1 if the river Type is Primary River and 0 if the river is a Tributary River.

Make a new DataFrame of just the Primary rivers.

Write a for loop that prints the name of each river in this new DataFrame.

Write another for loop that only prints the name of the river if it starts with the letter M.

Modify the for loop so it saves the names of these rivers as a list.

**a. Add a column called Primary that has value of 1 if the river Type is Primary River and 0 if the river is a Tributary River.**

```
In [13]: # Create a DataFrame to store the information
ColumnNames = ['Primary']
PrimeOrTributary = pd.DataFrame(0, index=np.arange(len(file)), columns=ColumnNames)

# Check if 'Primary' column already exists in file DataFrame
if 'Primary' not in file.columns:
    # Loop through each row in the file DataFrame
    for index, row in file.iterrows():
        # Check if the river is a primary river
        if row['Type'] == 'Primary river':
            # If it is, set the value to 1 in the PrimeOrTributary DataFrame
            PrimeOrTributary.loc[index, 'Primary'] = 1
        else:
            # Otherwise, set the value to 0
            PrimeOrTributary.loc[index, 'Primary'] = 0

    # Concatenate the file DataFrame with the PrimeOrTributary DataFrame
    file = pd.concat([file, PrimeOrTributary], axis=1)

# Print the updated DataFrame to verify the result
print(file)
```

	River	Continent	Average discharge (m3/s)	Length (km)	\
0	Amazon	South America	224000	6992	
1	Congo	Africa	41400	4370	
2	Nile	Africa	2810	6853	
3	Mississippi	North America	21300	3766	
4	Rio de la Plata	South America	27225	290	
5	Ob	Asia	13100	4345	
6	Parana	South America	19706	4880	
7	Yenisei	Asia	20200	4090	
8	Lena	Asia	18300	4294	
9	Niger	Africa	7900	4200	
10	Ganges	Asia	43950	3969	
11	Amur	Asia	11526	4444	
12	Yangtze	Asia	31900	6300	
13	Mackenzie	North America	9800	4241	
14	Irtys	Asia	2980	4248	
15	Volga	Europe	8220	3531	
16	Madeira	South America	30173	3380	
17	Missouri	North America	2445	3767	
18	Saint Lawrence	North America	16800	500	
19	Zambezi	Africa	4217	2574	
20	Ganges	Asia	18691	2525	
21	Paraguay	South America	4696	2621	

	Drainage area (km2)	Outflow	Type	Primary
0	6915000	Atlantic Ocean	Primary river	1
1	3822000	Atlantic Ocean	Primary river	1
2	3349000	Mediterranean Sea	Primary river	1
3	3248000	Gulf of Mexico	Primary river	1
4	3182064	Atlantic Ocean	Primary river	1
5	2990000	Gulf of Ob	Primary river	1
6	2582672	Rio de la Plata	Tributary river	0
7	2580000	Kara Sea	Primary river	1
8	2490000	Laptev Sea	Primary river	1
9	2117000	Gulf of Guinea	Primary river	1
10	1999000	Bay of Bengal	Primary river	1
11	1855000	Sea of Okhotsk	Primary river	1
12	1840000	East China Sea	Primary river	1
13	1805200	Beaufort Sea	Primary river	1
14	1673470	Ob	Tributary river	0
15	1380000	Caspian Sea	Primary river	1
16	1376000	Amazon	Tributary river	0
17	1371010	Mississippi	Tributary river	0
18	1344200	Gulf of Saint Lawrence	Primary river	1
19	1331000	Indian Ocean	Primary river	1
20	1300000	Bay of Bengal	Primary river	1
21	1120154	Parana	Tributary river	0

## b. Make a new DataFrame of just the Primary rivers.

```
In [14]: # Make a dataframe where the values within the 'Type' Column is 'Primary'
PrimaryRiverDF = file[file['Type'] == 'Primary river'].reset_index(drop=True) # we
print(PrimaryRiverDF)
```

	River	Continent	Average discharge (m3/s)	Length (km)	\
0	Amazon	South America	224000	6992	
1	Congo	Africa	41400	4370	
2	Nile	Africa	2810	6853	
3	Mississippi	North America	21300	3766	
4	Rio de la Plata	South America	27225	290	
5	Ob	Asia	13100	4345	
6	Yenisei	Asia	20200	4090	
7	Lena	Asia	18300	4294	
8	Niger	Africa	7900	4200	
9	Ganges	Asia	43950	3969	
10	Amur	Asia	11526	4444	
11	Yangtze	Asia	31900	6300	
12	Mackenzie	North America	9800	4241	
13	Volga	Europe	8220	3531	
14	Saint Lawrence	North America	16800	500	
15	Zambezi	Africa	4217	2574	
16	Ganges	Asia	18691	2525	

	Drainage area (km2)	Outflow	Type	Primary
0	6915000	Atlantic Ocean	Primary river	1
1	3822000	Atlantic Ocean	Primary river	1
2	3349000	Mediterranean Sea	Primary river	1
3	3248000	Gulf of Mexico	Primary river	1
4	3182064	Atlantic Ocean	Primary river	1
5	2990000	Gulf of Ob	Primary river	1
6	2580000	Kara Sea	Primary river	1
7	2490000	Laptev Sea	Primary river	1
8	2117000	Gulf of Guinea	Primary river	1
9	1999000	Bay of Bengal	Primary river	1
10	1855000	Sea of Okhotsk	Primary river	1
11	1840000	East China Sea	Primary river	1
12	1805200	Beaufort Sea	Primary river	1
13	1380000	Caspian Sea	Primary river	1
14	1344200	Gulf of Saint Lawrence	Primary river	1
15	1331000	Indian Ocean	Primary river	1
16	1300000	Bay of Bengal	Primary river	1

c. Write a for loop that prints the name of each river in this new DataFrame.

```
In [15]: # for loop, iterates through each row in the new dataframe 'PrimaryRiverDF'
for index, row in PrimaryRiverDF.iterrows():
    print(row['River']) # Look at the Rows under the 'River' header and print them
```



Amazon  
 Congo  
 Nile  
 Mississippi  
 Rio de la Plata  
 Ob  
 Yenisei  
 Lena  
 Niger  
 Ganges  
 Amur  
 Yangtze  
 Mackenzie  
 Volga  
 Saint Lawrence  
 Zambezi  
 Ganges

**d. Write another for loop that only prints the name of the river if it starts with the letter M.**

```
In [16]: for index, row in PrimaryRiverDF.iterrows():
         if row['River'].startswith('M'):
             print(row['River'])
```

Mississippi  
 Mackenzie

**e. Modify the for loop so it saves the names of these rivers as a list.**

```
In [17]: M_rivers = [] # Create an empty list

         for index, row in PrimaryRiverDF.iterrows(): # index each row within 'PrimaryRiverD
             if row['River'].startswith('M'): # If name of variable at index of row, under '
                 M_rivers.append(row['River']) # Append the variable at that index from 'row

         print(M_rivers)
```

['Mississippi', 'Mackenzie']