ANSWERS:

Q1:

*N*=8, *M*=24

mean: 1859.932954

sd: 855.0698853

*N*=50, *M*=150

Counter: 240000000

Number of Valid Entries 935314

Last Entry: 96745881600000000000

Numbers: [5, 3, 2, 1, 2, 1, 2, 2, 1, 4, 5, 1, 5, 4, 5, 1, 4, 1, 1, 5, 3, 2, 2, 2, 5, 2, 1, 6, 2, 3, 4, 5, 2, 3, 5, 6, 2, 6, 3, 1, 1, 2, 5, 3, 5, 5, 1, 1, 3, 4]

Mean: 1.864076520e+20

SD: 2.803021644e+20

Q2

median trip duration, in seconds: 629.0

What fraction of rides start and end at the same station: 0.02235839134

What is the standard deviation of the number of stations visited by a bike: 54.54189653

What is the average length, in kilometers, of a trip: 1.760204920

Calculate the average duration of trips for each month in the year. (Consider a trip to occur in the month in which it starts.) What is the difference, in seconds, between the longest and shortest average durations: 430.5702960

Let us define the hourly usage fraction of a station to be the fraction of all rides starting at that station that leave during a specific hour. A station has surprising usage patterns if it has an hourly usage fraction for an hour significantly different from the corresponding hourly usage fraction of the system as a whole. What is the largest ratio of station hourly usage fraction to system hourly usage fraction (hence corresponding to the most "surprising" station-hour pair): 1405.548725 37

There are two types of riders: "Customers" and "Subscribers." Customers buy a short-time pass which allows 30-minute rides. Subscribers buy yearly passes that allow 45-minute rides. What fraction of rides exceed their corresponding time limit: 0.03810678017

Most of the time, a bike will begin a trip at the same station where its previous trip ended. Sometimes a bike will be moved by the program, either for maintenance or to rebalance the distribution of bikes. What is the average number of times a bike is moved during this period, as detected by seeing if it starts at a different station than where the previous ride ended: 1038.730329

Q3

**Propose a project.\***

The project is to investigate the drying Aral Sea, which is located in Central Asia, between Kazakshtan and Uzbekistan. The goal is to analyze the main causes of the ecological disaster, its effect on health condition of local population, effectiveness of governmental policies to restore Aral Sea to its original size and predict future trend and its impact to local population.

Here are some facts about Aral Sea from Wikipedia:

The **Aral Sea** was an [endorheic lake](https://en.wikipedia.org/wiki/Endorheic_basin) lying between [Kazakhstan](https://en.wikipedia.org/wiki/Kazakhstan) in the north and [Uzbekistan](https://en.wikipedia.org/wiki/Uzbekistan) in the south.

Formerly one of the four [largest lakes](https://en.wikipedia.org/wiki/Largest_lakes) in the world with an area of 68,000 km2 (26,300 sq mi), the Aral Sea has been steadily shrinking since the 1960s after the rivers that fed it were diverted by [Soviet](https://en.wikipedia.org/wiki/Soviet_Union) irrigation projects. By 2007, it had declined to 10% of its original size, splitting into four lakes. Satellite images taken by NASA in August 2014 revealed that for the first time in modern history the eastern basin of the Aral Sea had completely dried up.

**Link to public description of data source.\***

Data was scrapped from CaWater info website. Here is the link for the data: <http://www.cawater-info.net/aral/data/morpho_e.htm>

Data consists of 88 observation points, one for each year between 1911 and 2006. There are 4 columns in years from 1911 to 1985 and 7 columns in years from 1986 to 2006. Dataset represents year, water level, water surface are and volume of Aral Sea. In latter years the sea was split into two parts, so data for these two parts of Aral Sea are given separately. It is also planned to search for other datasets related to health conditions, birth and death rate of local population during the project.

<http://tamirlan1.tumblr.com/post/152535000078/the-data-incubator-challenge-data-description>

<http://tamirlan1.tumblr.com/post/152535165298>

<http://tamirlan1.tumblr.com/post/152535171523>

10 (planning to analyze more)

<https://www.youtube.com/embed/ymQL9N0xyqc>