CSC 143 Assignment #2
Testing Benford's Law
For 25 points
Due Date: see Canvas class site

The Case Study described in Section 7.6 of our text book describes Benford's Law:

> Benford's Law involves looking at the first digit of a series of numbers. For example, suppose that you were to use a random number generator to generate integers in the range of 100 to 999 and you looked at how often the number begins with 1, how often it begins with 2, and so on. Any decent random number generator would spread the answers out evenly among the 9 regions, so we'd expect to see each digit about one-ninth of the time (11.1%). But with a lot of real-world data, we see a very different distribution.

Table 7.3 gives the expected distribution under Benford's Law:

1: 30.1%, 2: 17.6%, 3: 12.5%, 4: 9.7%, 5: 7.9%, 6: 6.7%, 7: 5.8%, 8: 5.1%, 9: 4.6%

You should use this distribution in your program.

You are being provided with a text file giving population data for 247 countries and territories. Your program should read in the data in this file and extract the information needed to keep count of how often each first digit is found. It may be useful to know that the ASCII value for the character '1' is 49.

In addition, you should analyze other publicly available data to see how they test for Benford analysis. One data repository site where you can find multiple datasets is: https://archive.ics.uci.edu/ml/datasets.html

For this assignment, you are to design a Java class named Benford that uses an *ArrayList* object to store counts of first digits. A Benford object will have only one field, the *ArrayList* object.

You will need the following class methods (at a minimum – you may add as many helper methods as you wish):

- A constructor to construct Benford objects
- A class method named *readCounts()* that takes a file name as a **parameter** and that reads and stores data from a text file into a Benford object.
- A class method named *benfordPercents()* that uses the data in the Benford object to fill an array of double values giving the percentage counts for each initial digit based on the raw counts contained in the Benford object.

You are also to design a client class named `BenfordPlot` that displays the data in the `Benford` object graphically. The output should be similar to the examples shown below, except your plot should be titled "Population of Countries." The "^" symbol in the plot below indicated the expected Benford value. The second plot shown below displays a distribution that does NOT follow Benford's Law.
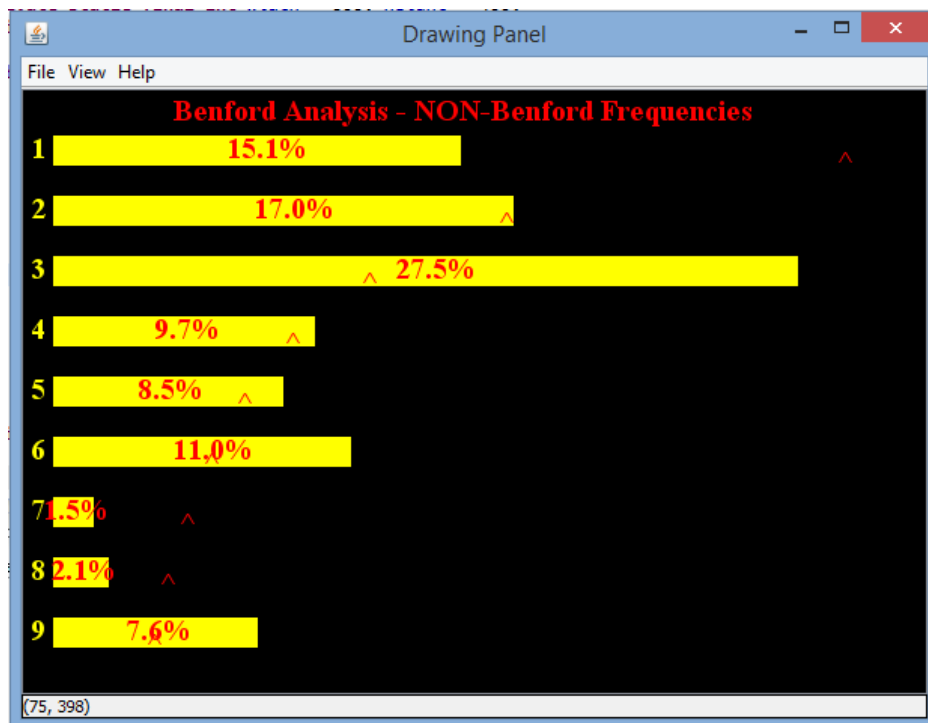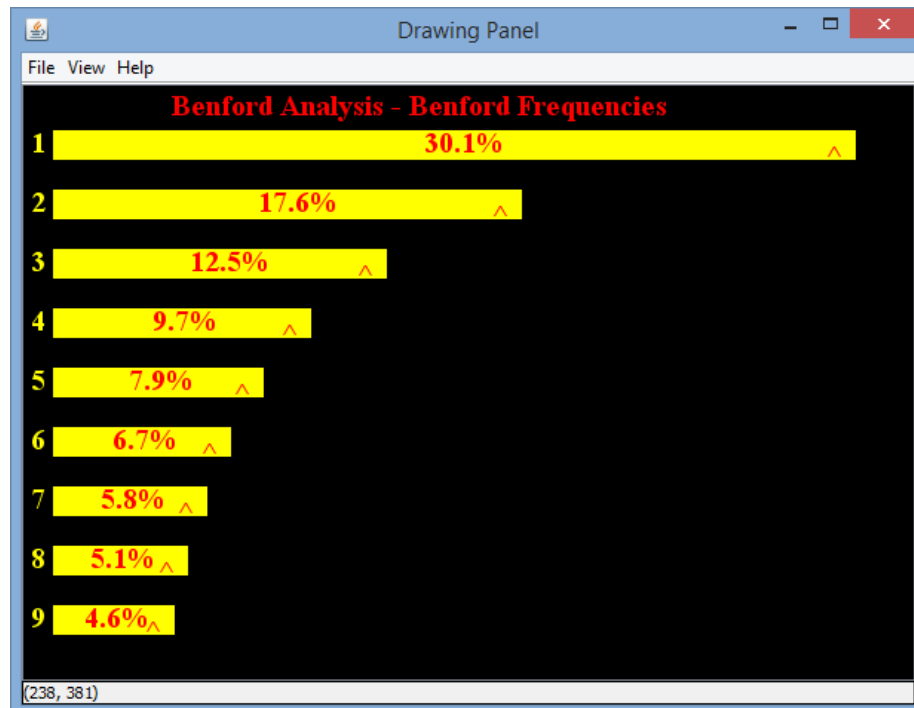
The file containing the data you need for this assignment is "popData.txt" – it was obtained from the website http://en.wikipedia.org/wiki/List_of_countries_and_dependencies_by_population

A sample of the data file you will be working with is shown below, so that you can see the format of the data you'll be using:

```
14   Ethiopia    90076012
15   Egypt   88404300
16   Germany 81083600
17   Iran    78287900
18   Turkey  77695904
19   Democratic Republic of the Congo    71246000
20   France  66121000
```

The fact that only the population data is numerical may help you parse the data file. Plot the population data provided to determine whether or not it follows Benford's Law.

The DrawingPanel.java code is also being provided, in case you do not already have it downloaded.

Please submit the following files:

- Benford.java (your Benford class code) - one per team.
- BenfordPlot.java (your client code that using the DrawingPanel code to produce a graphical display of the data contained in a Benford object) – one per team
- Analysis of multiple data sets along with your opinion as to whether or not the data follow Benford's Law – one per person

Design and implementation guidelines:

- Javadoc comment all class files and methods
- Use validation and generate and handle exceptions as appropriate
- Structured code - use methods to eliminate redundancy and break large methods into smaller, logical subproblems