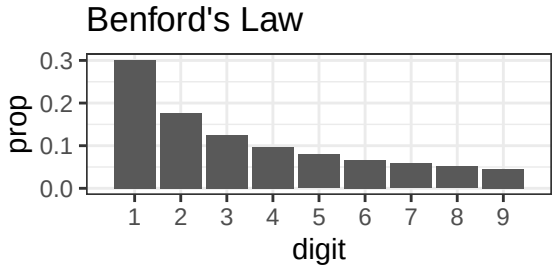
Lab 5: Informal Inference



Part I: 2009 Iran Election

On June 12 2009, the Republic of Iran held an election where President Mahmoud Ahmadinejad sought re-election against three challengers. One of the challengers, Mir-Hossein Mousavi. When it was announced that Ahmadinejad had won handily, there were widespread allegations of election fraud. There are many methods, both quantitative and qualitative, to detect election fraud. In this lab we will explore just one proposed method.

```
library(tidyverse)
library(stat20data)
data(iran)
```

Exploratory Data Analysis

1. What is the unit of observation in the `iran` data frame? What are the dimensions?
2. Which cities had the highest proportion of total votes cast for Ahmadinejad? Please return the top several city names along with the province name and the proportions.
3. Which cities had the highest proportion of total votes cast for Mousavi? Please return the top several city names along with the province name and the proportions.
4. How many cities did Mousavi win?
5. *How does the proportion of total votes that were voided compare between cities won by Mousavi and cities won by Ahmadinejad? This can be answered either with a plot or with summary statistics. Describe in words how they compare.
6. *What proportion of the total votes cast were won by Ahmadinejad and Mousavi, respectively?

First Digit Distribution

7. What proportion of vote counts for Ahmadinejad have "1" as a first digit?
8. *Create a plot showing the distribution of first digits in the Ahmadinejad's vote counts. Does this plot appear to match the ideal Benford's distribution? Where does it deviate?
9. *Would you consider this meaningless, weak, moderate, or strong evidence of election fraud? Why or why not?

Part II: Released before Tuesday Lab