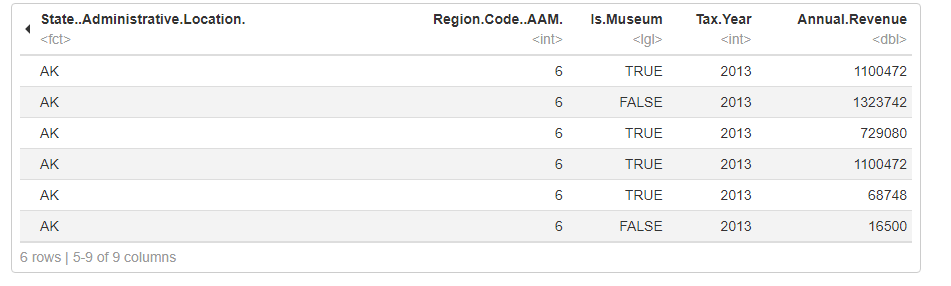
**The results of the project**

We have administrative records from the Institute of Museum and Library Services, IRS records, and private foundation grantmaking records. This dataset reflects the name, museums types, their classifications, their location and their annual revenues.

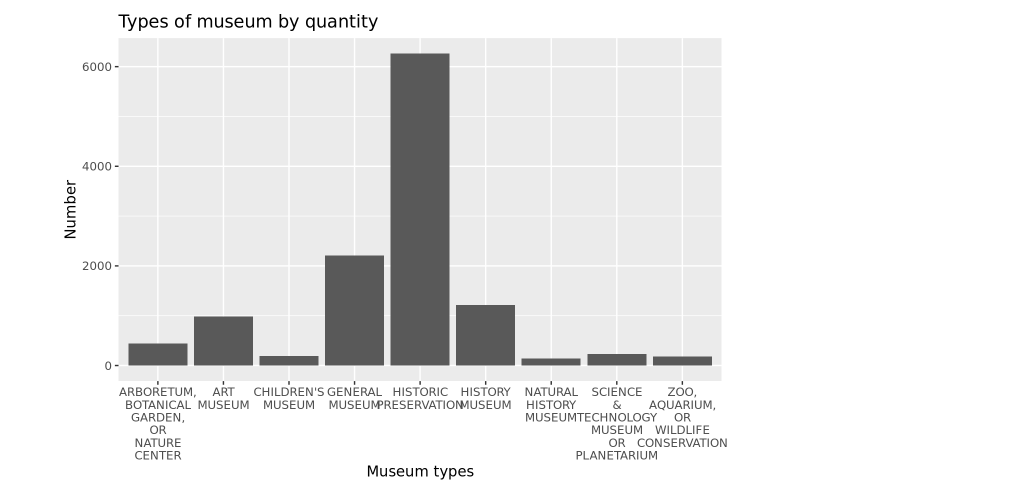
+ At first, we need to import our dataset into R and begin to inspect our data. We only focus on 6 aspects.



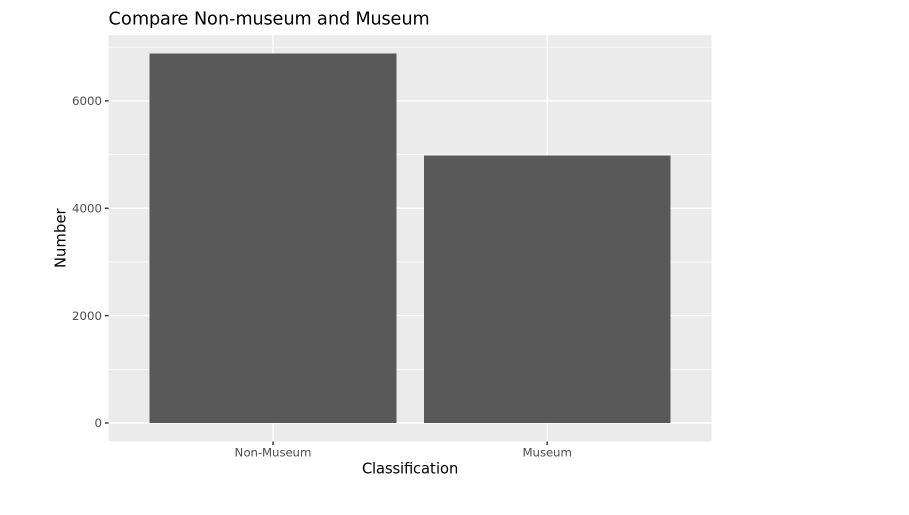
* Museum.Name: represents the name of each individual institution
* Museum.Type: Desribes what kind of museum each location is: a zoo, history museum, an aquarium,…
* Is.Museum: Classifies whether each institution will be called museum or not. For ex: Zoos, aquariums, historic sites aren’t really museums.
* Annual.Revenue: Reveals the income of each institution annually.
* State..Administrative.Location: Tells us the states where each museum locates.
* Region.Code..AAM: Assigns each state to the region code.

1. Explore institution by types

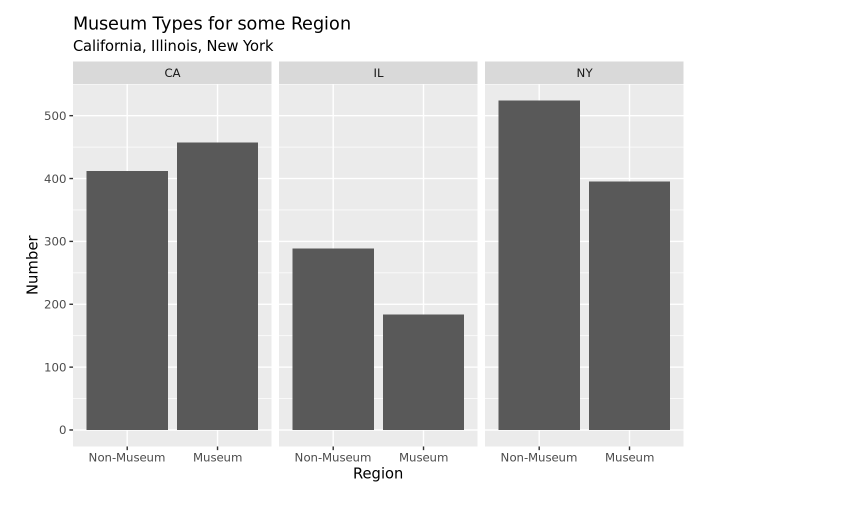
+ Second, we want to inspect the distribution of museums types.



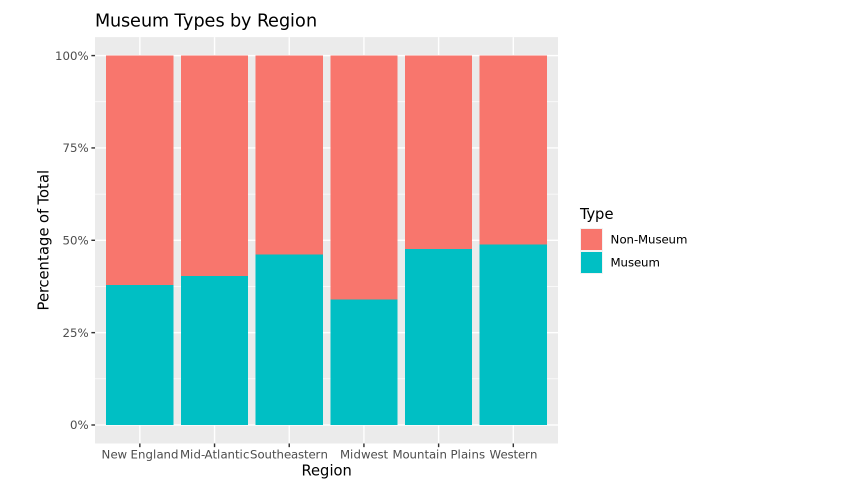
+ Third, we want to classify and see the number between Non-Museum and Museum.



+ Fourth, we want to go deeper. We will inspect their classification and their location as well. Start with some familiar states such as California, Illinois and New York.



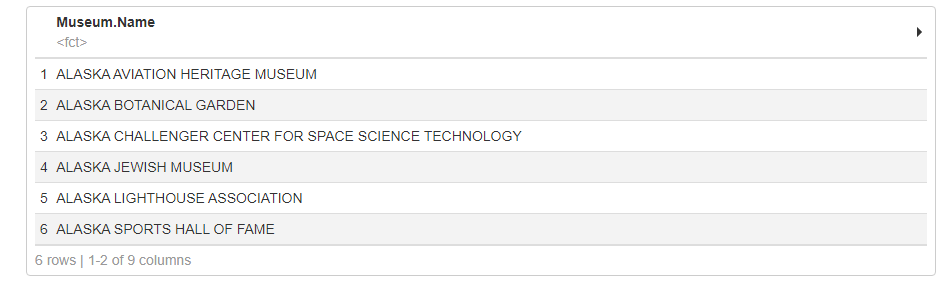
+ Fifth, we will start with the whole country.



2.Explore institutions by revenue

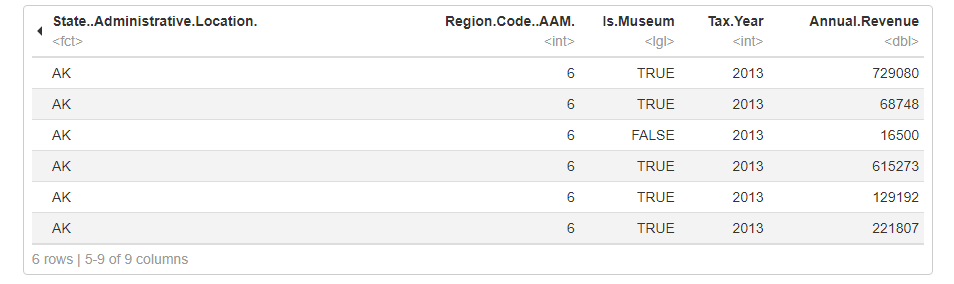
With this aspects, we’ll switch to looking at how much money each institution brought in and how that varies across geographies. Because we only have revenue data at the parent organization level, we’ll want to first filter our dataset to omit any duplicates. Next, we’ll create a few data frames from our starting data to look at different groups of museums by how much money they bring in.

+ Sixth, let filter our data.

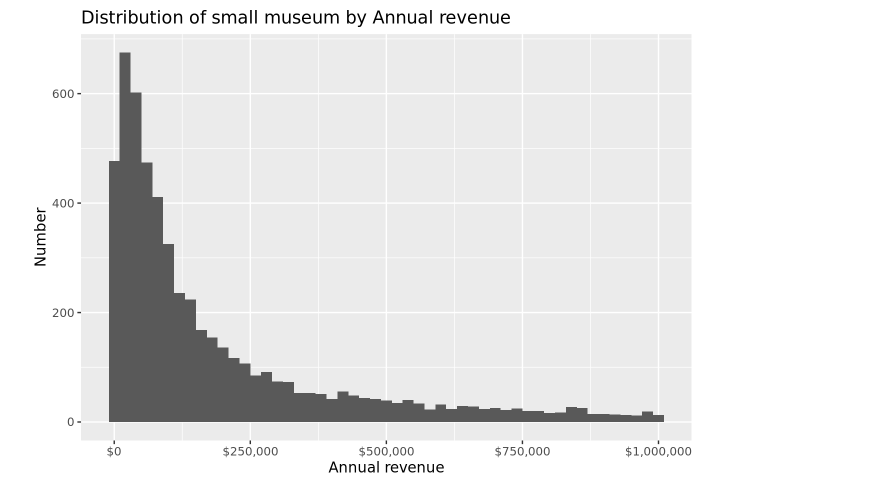


+ Seventh, we will visualize the distribution of small museums by Annual revenue.

* First, we need to filter our data based on the criteria: The Annual revenue must lower than 1,000,000 dollars. We must create new dataset for this filtered data so that it won’t overwrite on our initial dataset.

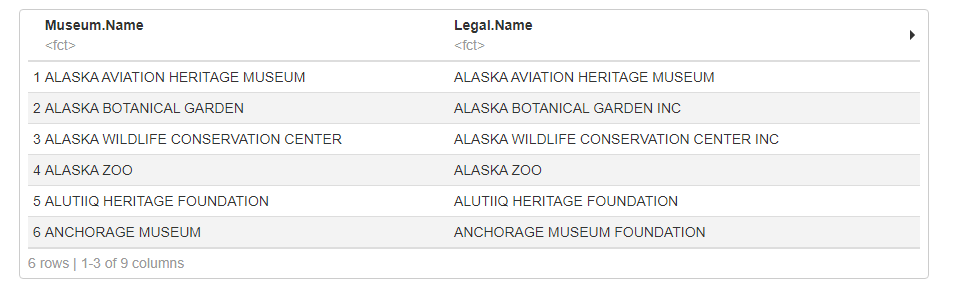


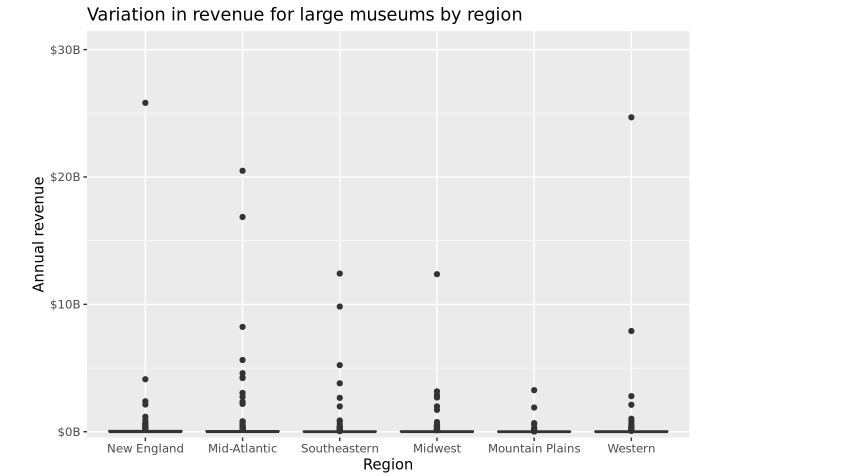
* Then we will use this new dataset to plot the graph. Best would be the histogram to show the distribution of a continuous variable.



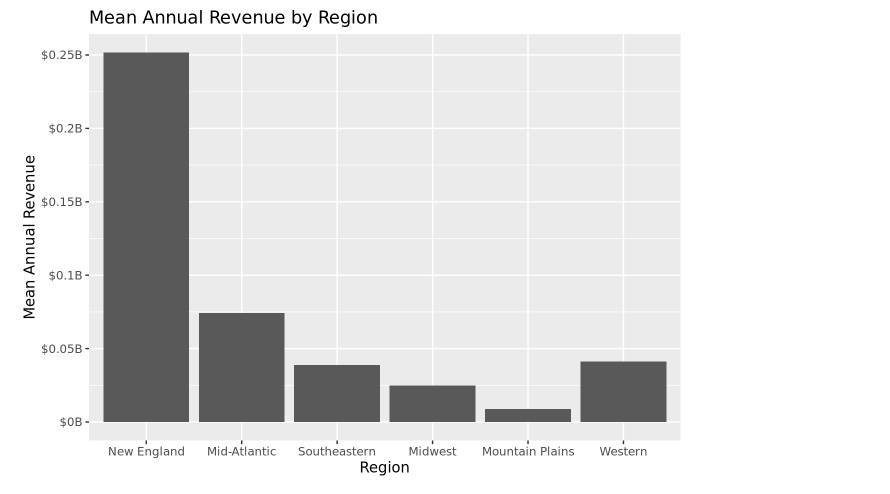
+ Eighth, we will look at the variation in revenue for large museums by region.

* Of course, we have to filter again and assign it to the new dataset.





+ Ninth, let’s take a look at revenue across all museums in our data by region.



+ Tenth, so when we have mean, it’s better to have error bars also.

* So from the old dataset, we will use it to calculate the mean and standard errors.

