

## **State of the Indian Lok Sabha**

### **Problem Statement:**

The task is to use the data on the Members of Parliament available online in the 15th Lok Sabha to show a series of visualizations that capture the state of the Indian Lok Sabha.

The data is in Excel and contains several things such as the political party to which the MPs belong, their educational qualifications, age, and most interestingly their activity in parliament -- attendance, debates in which they participated, questions they raised in parliament, etc.

### **Input Specification:**

The data set on the MP's available online is the input data.

### **Output Specification:**

Following are the list of visualizations that we have built:

1. Online Visualizations:
  1. Bar/Column Charts
  2. Bubble Chart
  3. Pie Chart
2. Using Prefuse on Java
  1. Mashup : Combined state and political parties together on the same visualization
  2. StateMenu : For viewing the while of the database in an organized manner
  3. TreeMap : For viewing the state wise political party distribution

The online visualizations are dynamic in nature and allow the user to view different parameters matched against various other parameters. We have also given the user the freedom to change the range of the data on which they want to visualize and hence have allowed them to perform query on the database itself.

The java Prefuse visualizations on the other hand are for the user to have a quick glance over the database, to view the summary and also analyse various statistics involved with the database provided.

### **The HTML Pages:**

#### **Scatter Plot:**

The Scatter Plot HTML page generates a Scatter Plot With the variables to choose from Age, Attendance, and Debates etc. to visualise the trends in the data as a whole without aggregating on any parameter. The Scatter Plot Data Ranges can be manually modified using a slider to Select Ranges for the above parameter.

**Comparison:**

The Comparison HTML page generates Bar Charts, Column Charts or Pie Charts with the data Grouped by State, Political Parties and Educational Qualification. The Values from the Grouped data can be manually changed so the user can compare various States, Political Parties or Educational Qualification Grouped by States, Political Parties or Education Qualifications.

**Grouping:**

The Comparison HTML page generates Bar Charts or Column Charts with the aggregate data of Age, attendance, debates etc. Grouped by State, Political Parties and Educational Qualification. The Values from the Grouped data can be manually changed so the user can compare trends in various States, Political Parties or Educational Qualification.

**Aggregate:**

The Aggregate HTML page generates Bar Charts, Column Charts or Pie Charts with the data Grouped by State, Political Parties and Educational Qualification. The Values from the Grouped data can be manually changed so the user can visualise the data of various States, Political Parties or Educational Qualification Grouped by States, Political Parties or Educational Qualification.

**Histogram:**

The Histogram HTML generates Bar Charts, Column Charts or Bubble Charts with aggregated values of age, attendance, debates etc. by grouping the data based on State, Political Parties, and Educational Parties etc. This is used to visualize the data. The variables for the histogram can be manually selected.

**The Google Charts API:****Using a Google Spread sheet as a Data Source:**

Visualizations built on the Google Visualization API can use any accessible Google Spread sheet as a data source.

Google Spread sheets support the Google Visualization API query language for sorting and filtering data.

Any selection of cells within a Google Spread sheet is a valid data source for visualization.

Using Google Spread sheets as a data source, use the URL of the sheet and cells that hold the data that you want. Specify that URL as the data source for your visualization in your visualization's user interface.

## **Sending a Request**

The Query object is instantiated with the URL of the Data source. The URL indicates what data is being requested, in a syntax understood by that data source. Sending method as an optional second parameter in the Query object constructor can be specified. A Query language string to sort or filter the results, and then send the request can be optionally added. The query language is a SQL language variant. The query is sent, specifying a call back handler that will be called when the response is received.

## **Processing the Response**

The response handler function is when the request returns. The parameter passed in to your response handler function is of type `google.visualization.QueryResponse`. If the request is successful, the response contains a data table (class `google.visualization.DataTable`). If the request failed, the response contains information about the error, and no `DataTable`.

## **Controls and Dashboards:**

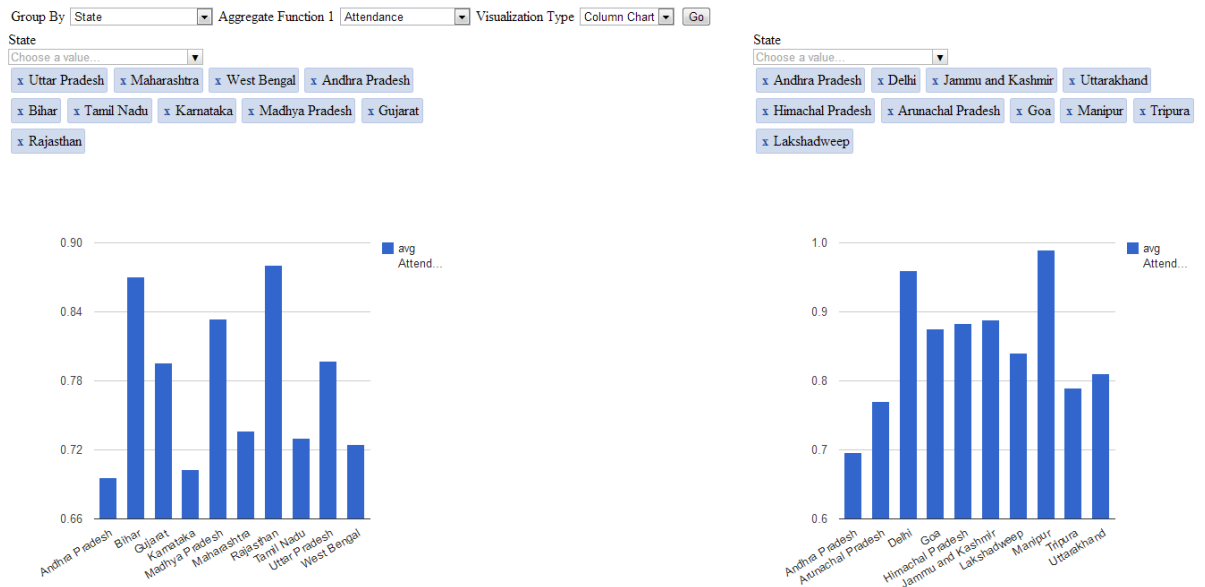
Dashboards provide a simple way to organize together and manage multiple charts that share the same underlying data.

Dashboard instances receive a `DataTable` containing the data to visualize and take care of drawing and distributing the data to all the charts that are part of the dashboard.

Controls are user interface widgets (category pickers, range sliders, autocompleters) user interacts with in order to drive the data managed by a dashboard and the charts that are part of it. `ControlWrapper` added to instances to a dashboard, where they behave like pipes and valves in a plumbing system. They collect user input and use the information to decide which of the data the dashboard is managing should be made available to the charts that are part of it.

# Hypothesis Testing

Hypothesis 1: States with larger number of constituencies have a lower average attendance in the Lok Sabha as compared with states with smaller number of constituencies.



Data from statistics:

The states were divided into 3 categories:

1. Large (Comprising of the top 10 states)
2. Medium (Comprising of next 8 states)
3. Small (comprising of the remaining states)

The following were the average attendances and the standard deviations across the groups:

Category	Average Attendance	Standard Deviation
Large	77.54%	16.12
Medium	74%	18.76
Small	86%	12.124

As we can see, the average age for small category of states is higher as compared to the other two categories, and has an even smaller value of standard deviation, so our hypothesis stands.

1. The average age of MP's in small states is higher as compared to the average age of MP's in the other two categories.

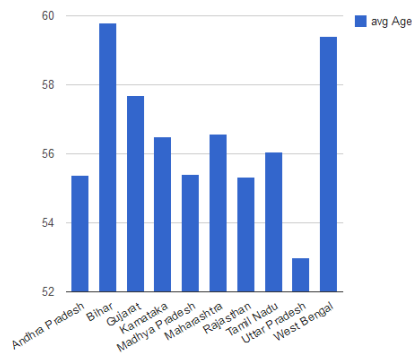
Group By:  Aggregate Function 1:  Visualization Type:

State  
Choose a value...

☒ Uttar Pradesh ☒ Maharashtra ☒ West Bengal ☒ Andhra Pradesh

☒ Bihar ☒ Tamil Nadu ☒ Karnataka ☒ Madhya Pradesh ☒ Gujarat

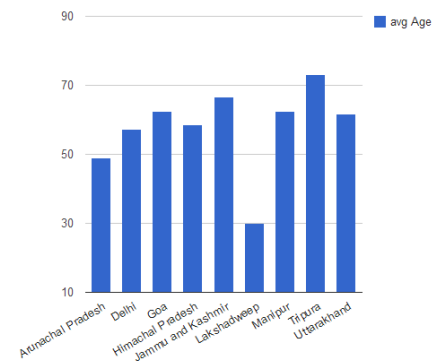
☒ Rajasthan



State  
Choose a value...

☒ Arunachal Pradesh ☒ Delhi ☒ Jammu and Kashmir ☒ Uttarakhand

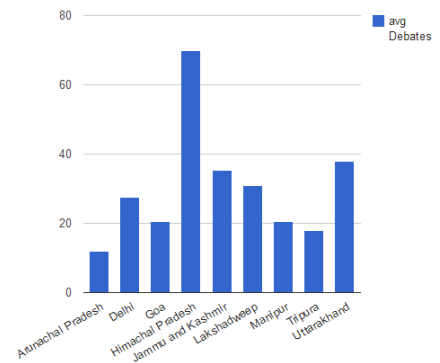
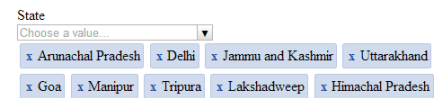
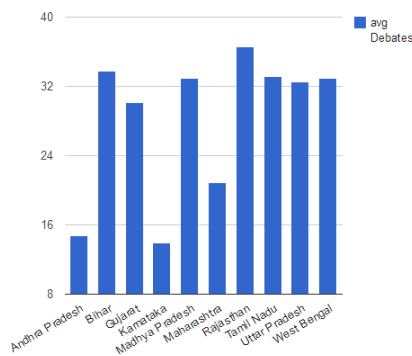
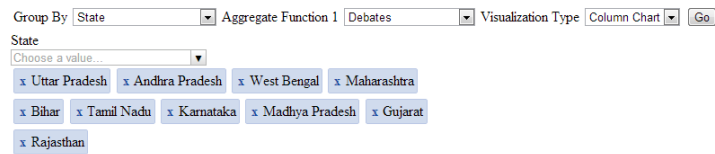
☒ Himachal Pradesh ☒ Goa ☒ Manipur ☒ Tripura ☒ Lakshadweep



Category	Average Age	Standard Deviation
Large	56.23	11.09
Medium	56.26	11.75
Small	59.34	11.66

The averages clearly indicate a trend among ages in the states differentiated on the basis of their size. Since the standard deviations are similar, we may conclude that the smaller states have a higher average age as compared with the other two categories. Our hypothesis stands.

2. Participation in the Lok Sabha (characterized by the number of debates) is higher for small sized states as compared with the other 2 categories.



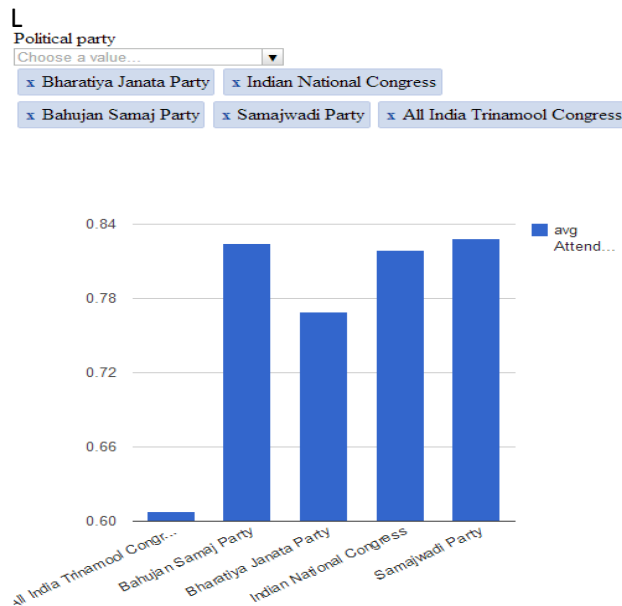
Category	Average Debates	Standard Deviation
Large	28.33	36.44
Medium	23.46	27.91
Small	30.77	29.34

One can see from the average participation in debates that the MP's from smaller state fare better than the medium state MP's. However, as compared with the large state MP's, the averages are close but the standard deviation presents another picture. By analysing the standard deviation values, one can say that participation in debates is in extremes across all the three categories, meaning that there are very few people actually participating in a large number of debates, while there are also people have extremely low participation in debates across all three categories.

Our hypothesis stands true.

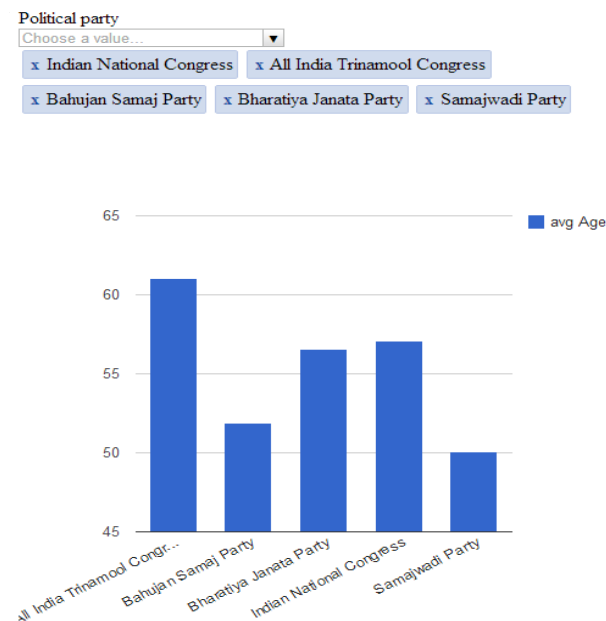
### 3. Comparing the Major Political Parties across our nation on the basis of Attendance

Hypothesis 4.1: Major Political parties have similar value of attendances.



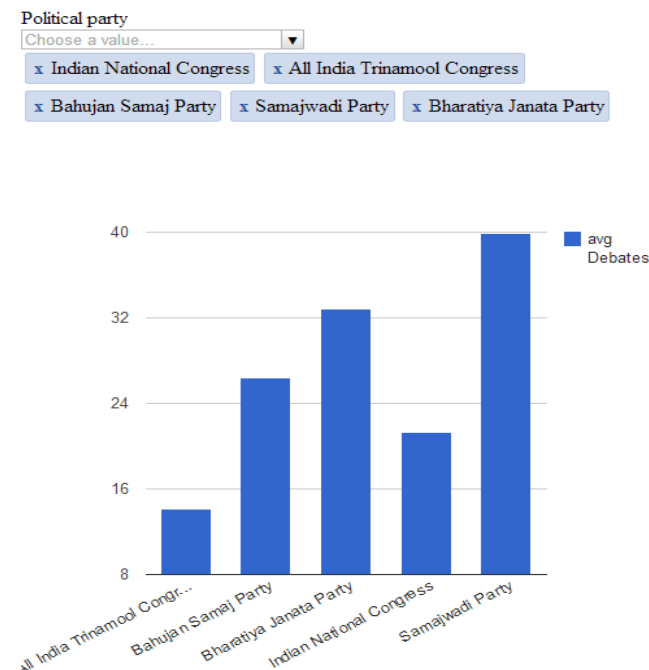
As one is able to see from the comparisons across the major political parties, the BJP, INC, BSP and Samajwadi Party have almost the average attendance in Lok Sabha. Surprising fact is the AITMC, one of the 5 major political parties in our country has a very low value of attendance as compared with the other 4. Let us have a look at whether this has some relation with the average age of these parties.

Hypothesis 4.2: Average Age of AITMC is higher than the other political parties (considering the old ones are the lazy ones)



Our hypothesis about the average age of AITMC turned out to be correct as shown by the visualization on the left. While the BJP and INC have an average age close to the overall averages (shows how they dictate even the averages in our political system), the samajwadi party and bahujan samaj party have lower average ages. Having looked at the ages, let us now also have a look at their participation in debates to see whether the young ones are also participating in debates.

Hypothesis 4.3: Parties with lower average age have a higher participation in debates.



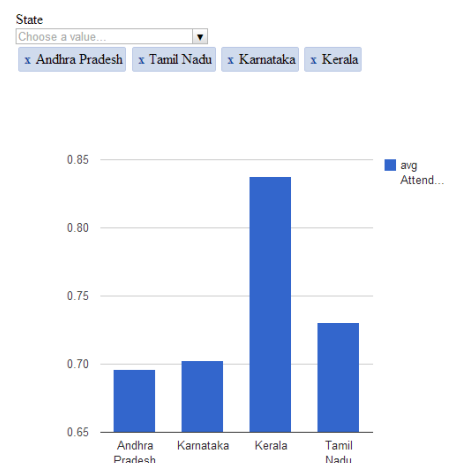
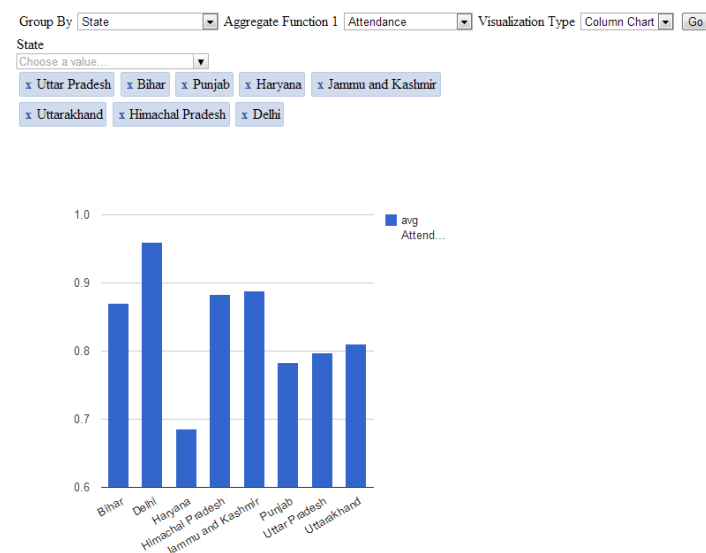
Looking at the averages, we may conclude that infact the AITMC with high average age, lower attendance has indeed a very low participation in debates. Compared with the Samajwadi Party on the other hand, who have a lower age and higher attendance, they have indeed a very active participation in debates. What we may conclude is that parties represented by young MP's are participating to a greater extent in our political system and they also have a higher attendance.

## The North-South divide

Hypothesis 5: Average age of MP's from North India is higher than average age of MP's from South India.

As we had done before, we partition the states into 5 different categories :

North, South, East, West and Central India



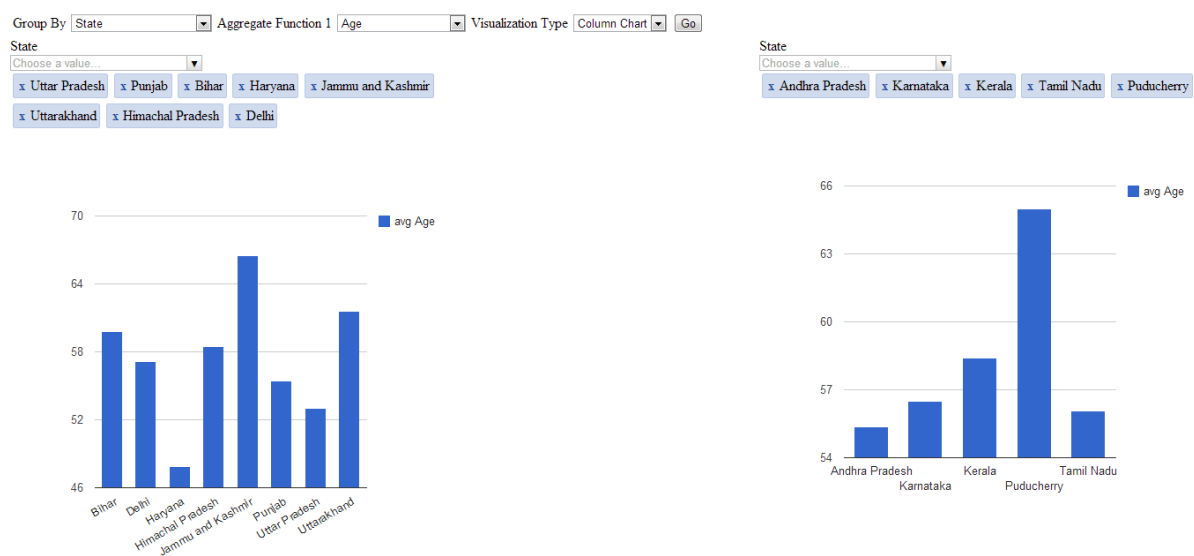
Now, let us have a look at the results from the statistics to analyse whether our hypothesis hold true or not.



Division	Average Attendance	Standard Deviation
Central	75%	19.23
East	77%	17.18
North	82%	14.48
South	73%	17.06
West	78%	16.12

Having analysed the spreadsheet statistically, we notice a trend that the average attendance of MP's from the North is higher as compared with the other categories. Also, the standard deviation of the North MP's is also low meaning that the MP's from the north are collectively attending the Lok Sabha sessions whereas MP's from the other zones are either attending most of the sessions or are attending very less sessions.

Hypothesis 6: Average Age of MP's from North is less than average age of MP's from South.

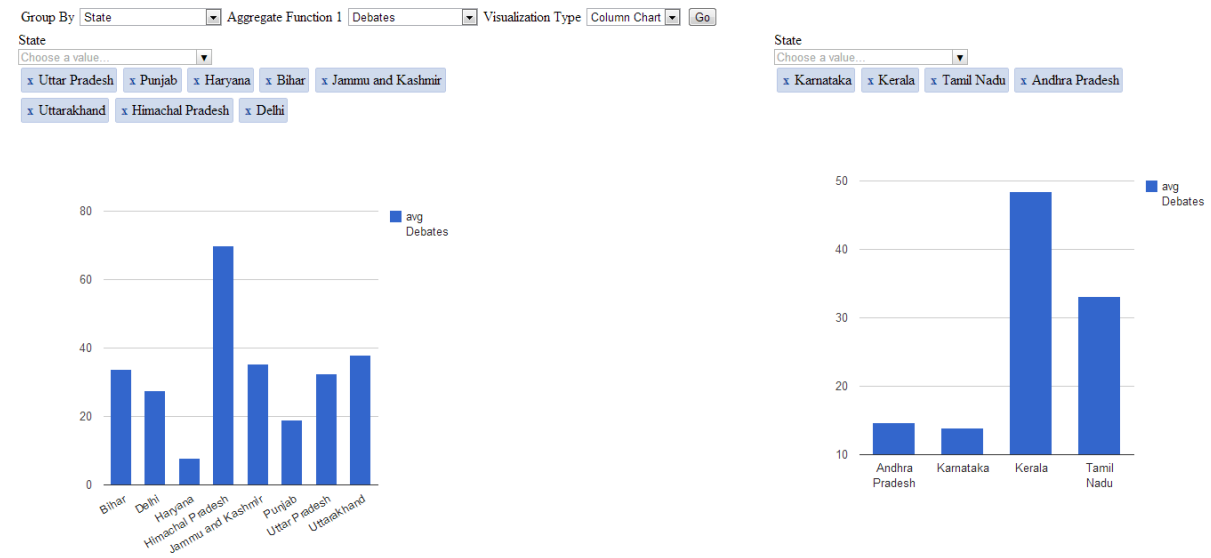


Division	Average Age	Standard Deviation
Central	55.43	11
East	60.12	11.34
North	55.65	12.12
South	56.39	9.7
West	56.54	11.39

On analysing the table above, we notice that MP's from South and North India have similar values of ages, but we notice through the standard deviation values that the MP's from South are more likely to belong to the 53-58 age category whereas there are a higher number of younger MP's from the North. An interesting trend to be noted here is the fact that MP's from East India have an abnormally high average age values. Maybe people on the Easter part of our nation place there trust on the old and wise.

Our hypothesis stands.

Hypothesis 7: Average Participation in Debates of MP's from South is less than that of MP's from North India.



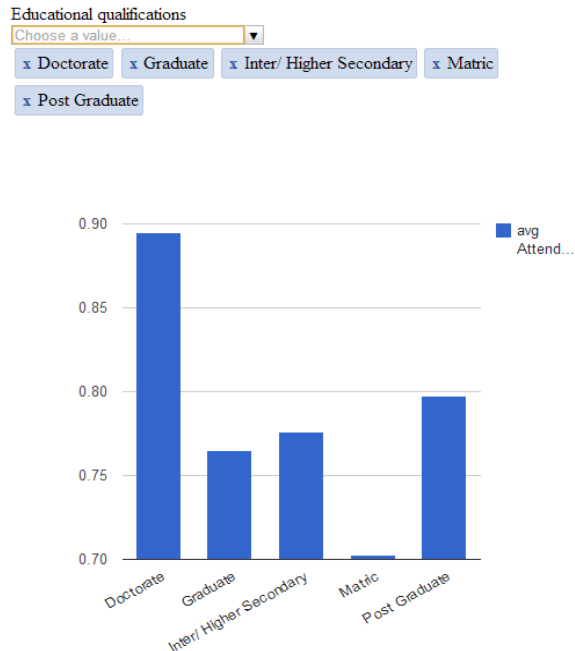
Division	Average Debates	Standard Deviation
Central	25.65	33.107
East	26.47	31.844
North	31.03	38.065
South	24.9	26.53
West	26.51	39.96

On analysing the table above, we see that the average participation in debates by MP's from north India is indeed higher as compared with MP's from the other division. Lower average age and higher attendance do indeed correlate with higher participation in debates.

Our hypothesis stands.

Comparing MP's with higher educational qualification with MP's with lower educational qualifications

Hypothesis 8.1 : MP's with higher educational qualifications have better attendance compared with others.

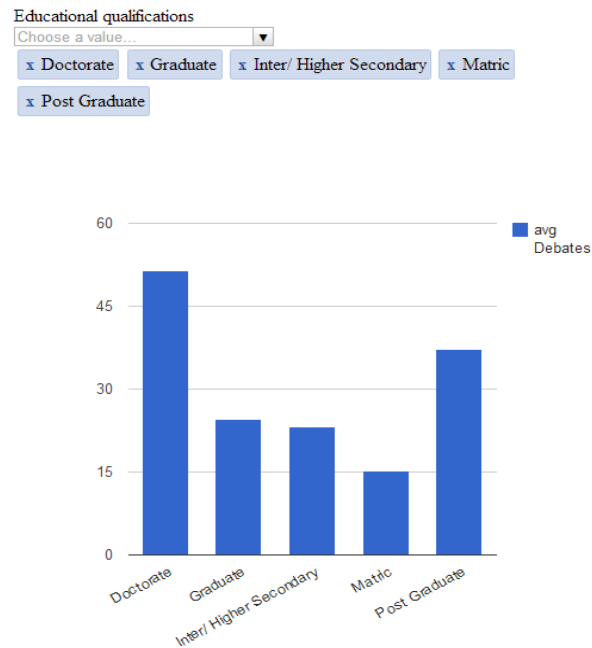


On having a look the visualization, we see that the MP's with doctorate and post-graduate degrees indeed have a higher attendance in the Lok Sabha. Maybe, they are the ones, with the highest years of schooling who realize the importance of attendance as compared with the others. The data from the sheet further supports this claim.

Hypothesis stands.

Division	Average Attendance	Standard Deviation
Doctorate	89%	7.6
Graduate	77%	16.4
Inter/higher secondary	78%	15
Matric	70%	21.4
Post graduate	80%	15.4

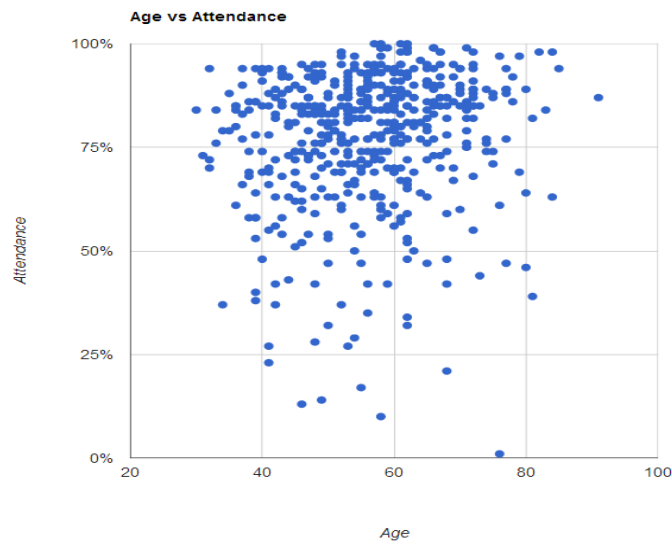
Hypothesis 8.2: MP's with higher educational qualifications have higher participation in debates as compared with others.



On having a look at the visualization and the related statistics from the table below, we see that the doctorates and the post graduates do in fact have a higher participation in debates as compared with the other categories . Higher education did give them a better understanding and are therefore more actively participating in debates in the Lok Sabha, MAYBE.

Division	Average Debates	Standard Deviation
Doctorate	51.4	50.32
Graduate	24.5	24.85
Inter/higher secondary	23.533	26.34
Matric	15.2	21.87
Post graduate	32.07	49.08

Hypothesis 9: Average attendance for MP's with age higher than 56 (mean age) is higher than their counter parts.



Null Hypothesis: Average attendance for MP's with (age > 56) = Average attendance for MP's with (age <= 56)

Degrees of Freedom: 548

T- Value: 3.54

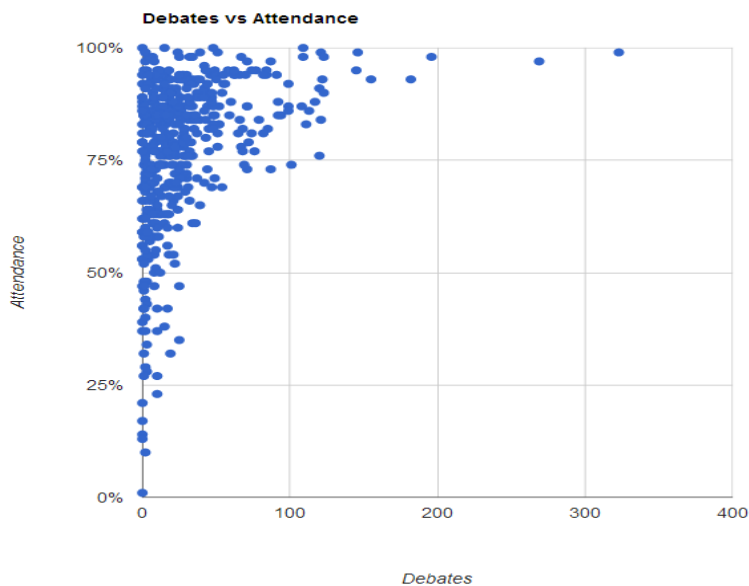
Critical Value: 1.960 (Significance Level = 5%)

T-value > Critical Value means that the null hypothesis is rejected and our hypothesis stands.

Having analysed the data above using Welch T-Test, we find that the older MP's are more actively involved (judging by their attendance levels) in the parliament as compared to the younger MP's.

Maybe the young ones have gone lazy over the past few years.

Hypothesis 10: Average debates for people having higher attendance are higher as compared to people with attendance lower than average (77%).



Null Hypothesis: Average debates for MP's with (attendance > 77) = Average debates for MP's with (attendance <=77).

Degrees of freedom: 449

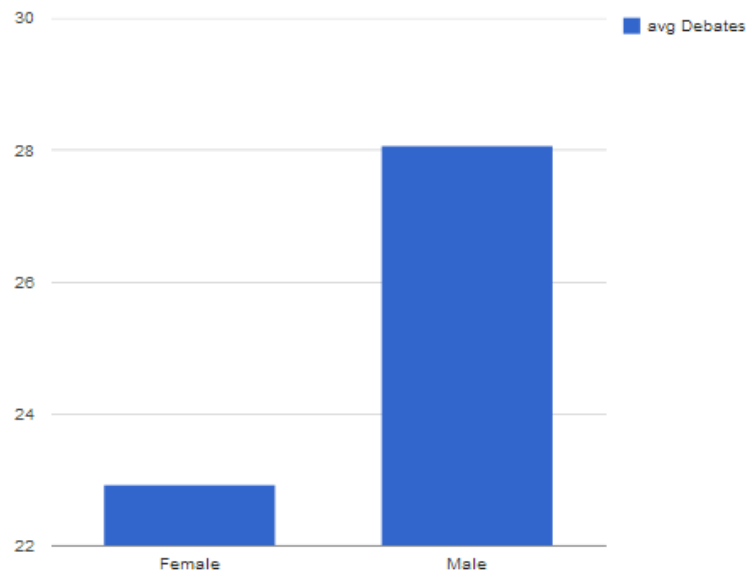
T-value: 7.70

Critical Value = 1.960 (Significance Level = 5%)

T-value > Critical Value means that the null hypothesis is rejected and our hypothesis stands.

This result was expected with our expectations that people who are attending the Lok sabha sessions more often are the ones more actively involved with the debates as well. The ones who are not attending the sessions are usually the inactive ones when it comes to debating issues of national importance.

Hypothesis 11: Males have a higher participation in debates as compared to females.



Null Hypothesis: Males and Females have equal participation in debates.

Degrees of freedom: 81

T-value: 1.94

Critical Value 1 = 1.960 (Significance Level = 5%)

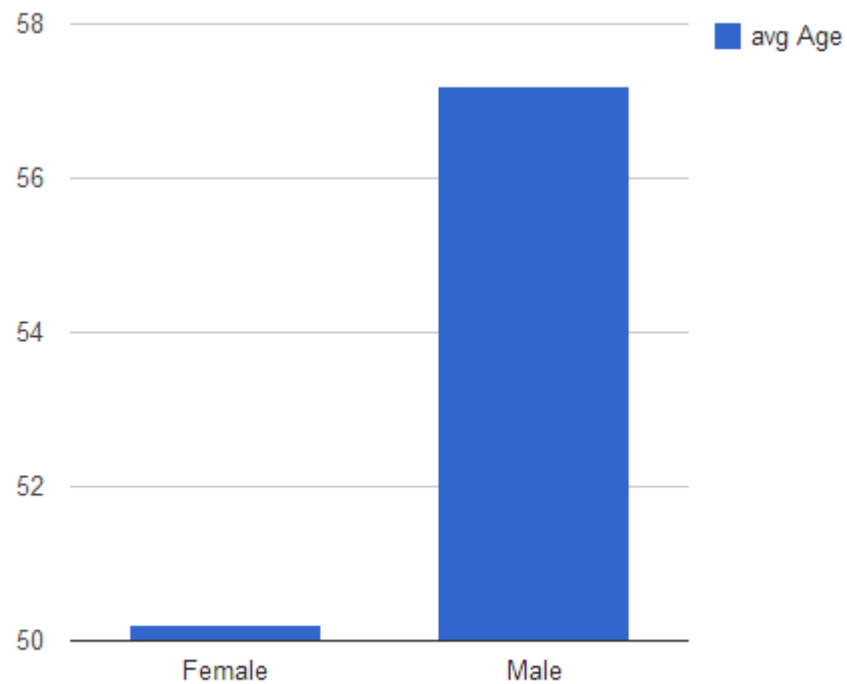
Critical Value 2 = 1.645 (Significance level = 10%)

T-value > Critical Value2 means that the null hypothesis is rejected with a confidence level of 90%.

However, if we are to increase the confidence level to 95%, then T-Value < Critical Value 1 and our null hypothesis stands.

Although the males are performing better in debates as compared to the average values, however on looking at the variances, we find that there are only a few Male and Female MP's who are actively involved in debates and thus comparison across the gender yields that their participation might be equal high a high chance.

Hypothesis 12: The average age of Male MP's is higher than the average age of female MP.



Null Hypothesis: Average age of Male MP's = Average age of Female MP's.

Degrees of freedom: 164

T-value: 1.744

Critical Value = 1.960 (Significance Level = 5%)

T-value < Critical Value means that the null hypothesis stands and that the male and females are of similar ages if the variances are also taken into account.