CSP301 Assignment 2

- -

State of the "15^{th} Indian Lok Sabha"

Abhishek Kumar (2011CS50272)
Akhil Jain (2011CS50274)
Shivanker Goel (2011CS10298)

03-Oct-2012

Contents

1 A hypothetical Social Networking Website

We were given live data from a hypothetical social media website and were asked to build a web based analytics dashboard to get a birds eye view of the activity on the website, such as the features provided by Google analytics. This hypothetical website has 2,500 users with approximately 65,000 edges between them. They come from interesting places such as Heaven and Hell and Asgard! And they don't sleep or eat or drink, they only gossip with each other on a wide variety of topics!

We were asked to show a series of visualizations capturing the essence of the 15^{th} Lok Sabha. Also, we were to carry out hypotheses testing in order to establish mathematically sound relationships between Members and their performance in the Parliament.

We were to submit a report consisting of all the interesting information that we could deduce from our analysis and visualizations.

2 Project Structure

A soft copy of the source code of our project is available at
https://github.com/CSDesign/furious-lion.git

We have created a dynamic Web User Interface using the **D3** functional programming toolkit of Javascript. All client side scripting has been done in javascript while the dynamization of the graphs has been done via server side scripting in **PHP**.

An instance of the Web UI may be run by opening the file index.php on our github repo (linked above).

For the purpose of analysis of the given data, we have carried out the **Welch's t test** between pairs of various parameters given in the data. Many interesting results were obtained, some of which are stated in the pages to follow.

In this project we have used 4 main languages – Javascript, HTML, CSS and PHP. We also used a lot of features of Dream Weaver, MS-Excel and LaTeX.

3 Data Analysis

Our data analysis is mainly divided into two parts - Hypothesis Testing and Visualisation based Analysis.
In order to make the analysis and various statistics calculated relevant, we have made the following assumptions:

1. Wherever applicable, the fields containing the value 'N/A' have been ignored.

2. For quantizing the educational qualifications of the various parliamentarians, we have assigned the following weightages to the qualifications:
   0 $\longrightarrow$ Under Matric
   1 $\longrightarrow$ Matric
   2 $\longrightarrow$ Inter/Higher Secondary, Certificate course
   3 $\longrightarrow$ Diploma, Under Graduate
   4 $\longrightarrow$ Graduate, Post Diploma Course, Professional Graduate
   5 $\longrightarrow$ Post Graduate
   6 $\longrightarrow$ Doctrate

3. For the purpose of qualifying as highly educated and otherwise, 0,1 and 2 have been classified as less educated while the others have been classified as highly educated.

3.1 Visualization

We tried to deduce many fun facts from what we visualized in the Project. Although not rigorously tested, the visualizations provided to us an intutive way to look at the data. It also showed us some important features about the data at hand which we have mentioned in the pages to follow. Our visualization mainly consisted of the following components:

1. Histograms
2. Pie Charts
3. Scatter Plots
4. Scatter Plot Matrix
5. Stacked Area Chart
6. Bar Charts
7. Statewise Distribution of MPs
8. State-Party Mashup

3.2 Hypothesis Testing

It is the process of testing to what extent do our hypotheses stand true when generalized to the population, and not limited by the sample at hand. In this project, we were mainly concerned with the relationships between various parameters that have been provided to us.

Central to every hypothesis test is the Null Hypothesis. Simply stated, the Null Hypothesis corresponds to their being no relation whatsoever between the parameters under study. The Alternate Hypothesis on the other hand asserts a particular relationship between the parameters.

In this project, Hypothesis Testing has been done using the **Welch's t test**.
In this test, we define the t test statistic as

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

where $\overline{X}_i$, s_i^2 and N_i are the i^{th} sample mean, sample variance and sample size, respectively. The degrees of freedom ν associated with this variance estimate is approximated by

$$\nu = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2 \cdot \nu_1} + \frac{s_2^4}{N_2^2 \cdot \nu_2}} = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2 \cdot (N_1 - 1)} + \frac{s_2^4}{N_2^2 \cdot (N_2 - 1)}}$$

Here $\nu_i = N_i - 1$, the degrees of freedom associated with the i^{th} variance estimate.

Once we obtain the values of t and ν, We can compare their values to the standard t-value distribution tables. For the purpose, we used the calculator on *http://www.danielsoper.com /statcalc3/calc.aspx?id=10*.

The null hypothesis is said to be true when $t_{obtained} < t_{critical}$. Otherwise, the alternate hypothesis is said to be true.

Now in order to decide which form of the Alternate Hypothesis is correct, we evaluate the mean value of the parameter under test for both the groups. The group with the higher mean value is declared to have swung the result in its favor. This leads to the conclusion that the group with the higher mean value is the winner of the t-test.

4 Visualization

4.1 Histograms

Age

We see that the distribution of ages in the Indian Parliament follows a pseudo-normal distribution. The youngest member of the Parliament is aged 30 years (Hamdullah Sayeed) while the oldest member is aged 91 years (Ram Sundar Das). The average age of the Parliament is 56 years. This shows that we have a wide spectrum of people from age groups present in the primary decision making and deliberating body of the country.

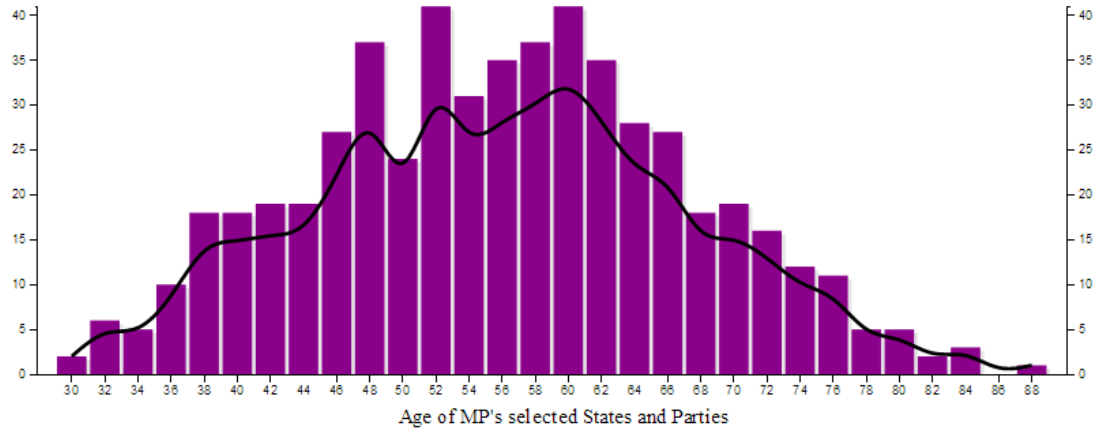

Fig 4.1.1: Age distribution in the Indian Parliament

Debates

We see that most people in the parliament prefer not to participate in debates at all or participate in very few debates. The national average for participation in Debates is 28.4. The most active parliamentarian in this regard is Arjun Ram Meghwal (BJP) with participation in 323 debates. This shows that in general the number of active MPs is very low.

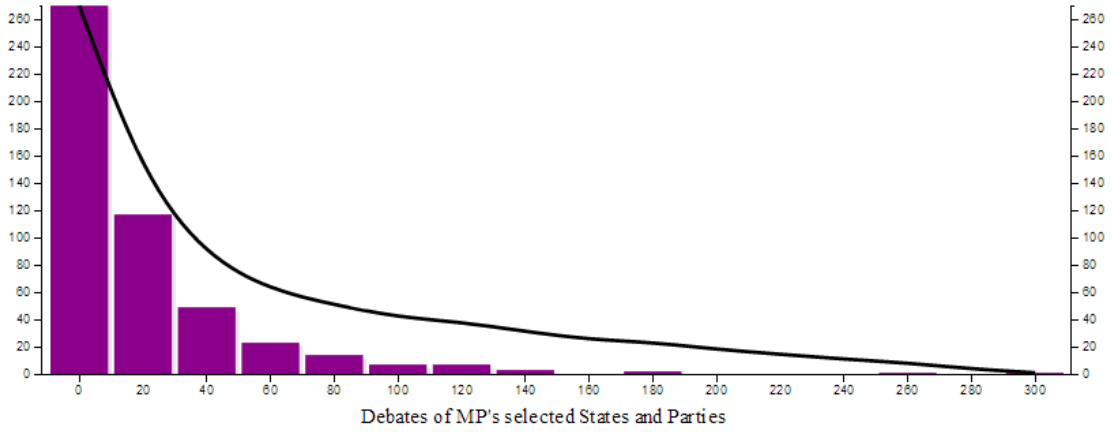

Fig 4.1.2: Debate participation in Indian Parliament

Questions

We see that most people are more active when it comes to asking questions in the parliament. An average MP asks 214 questions. The most active Parliamentarian in this regard is Anandrao Adsul (Shiv Sena) with 886 questions. However, most people still come in the bracket of no or few questions as is evident from the following graph:

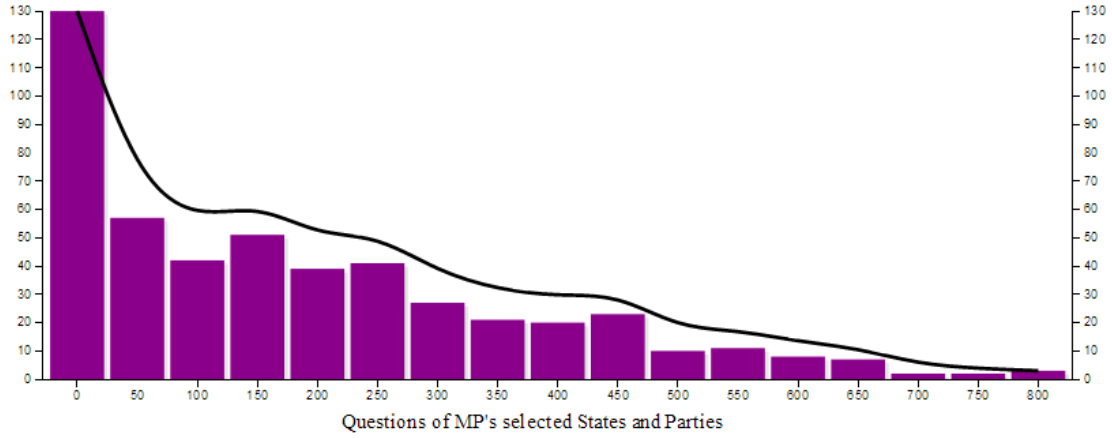

Fig 4.1.3: Questioning Patterns in Indian Parliament

The distribution of questions among the two major political alliances is as follows:

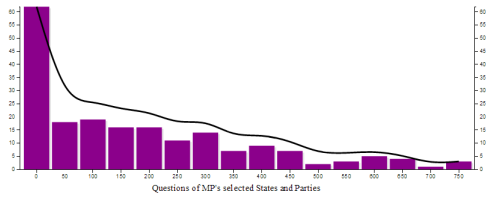

Fig 4.1.4: Questions asked by UPA

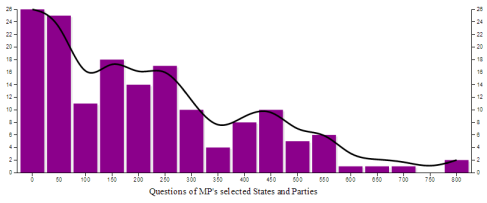

Fig 4.1.5: Questions asked by NDA

Thus it is seen that almost all members of the principal opposition ask questions while most members of the party in power do not.

Private Member Bills

It is seen that most members do not introduce Private Member Bills on the floor of the Parliament. However, the most active member in this regard is Hansraj Gangaram Ahir (BJP) with 24 Bills. A distribution of private member bills with MPs can be seen from the following graph:

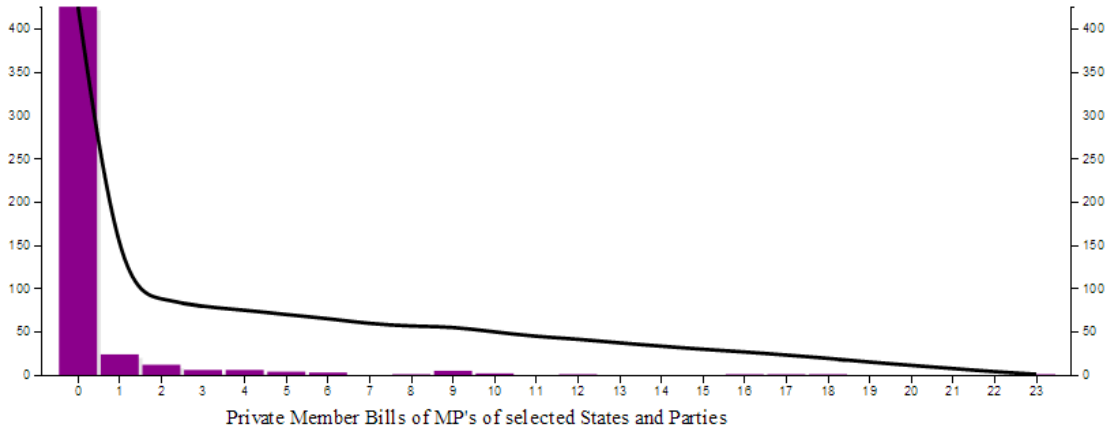

Fig 4.1.6: Private Member Bills in the Indian Parliament

Attendance

When we switch to attendance, we obtain a bell curve with the peak near an 85% value. The average attendance of the Parliament is 77%. This is more clearly visible from the following graph:

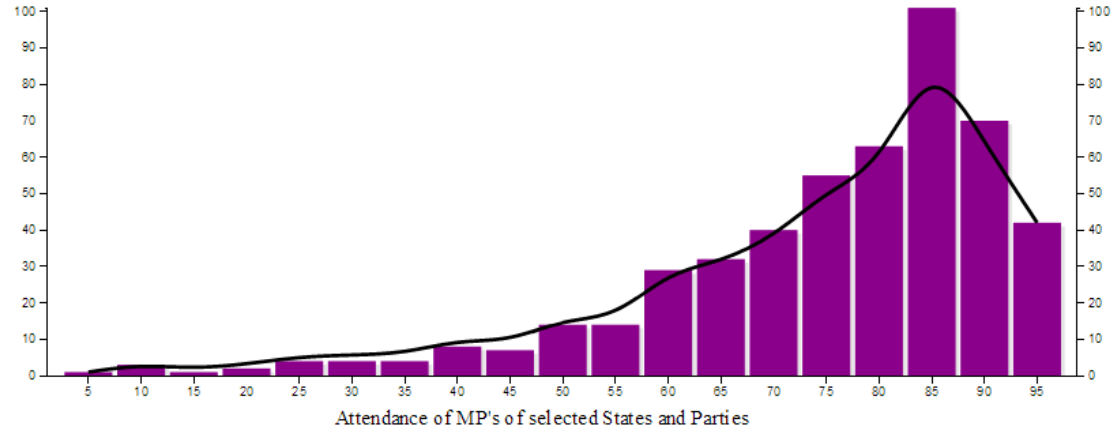

Fig 4.1.7: Attendance in the Parliament

The distribution of attendance among the two major political alliances is as follows:

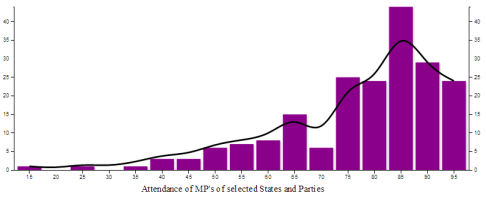

Fig 4.1.8: UPA Attendance

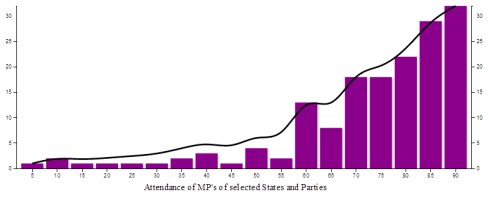

Fig 4.1.9: NDA Attendance

Thus it is seen that most members of the principal opposition maintain high attendance in the Parliament in comparison to the members of the party in power.

4.2 Pie Charts

Political Parties

The All India Distribution of political parties is as shown by the following graphic:

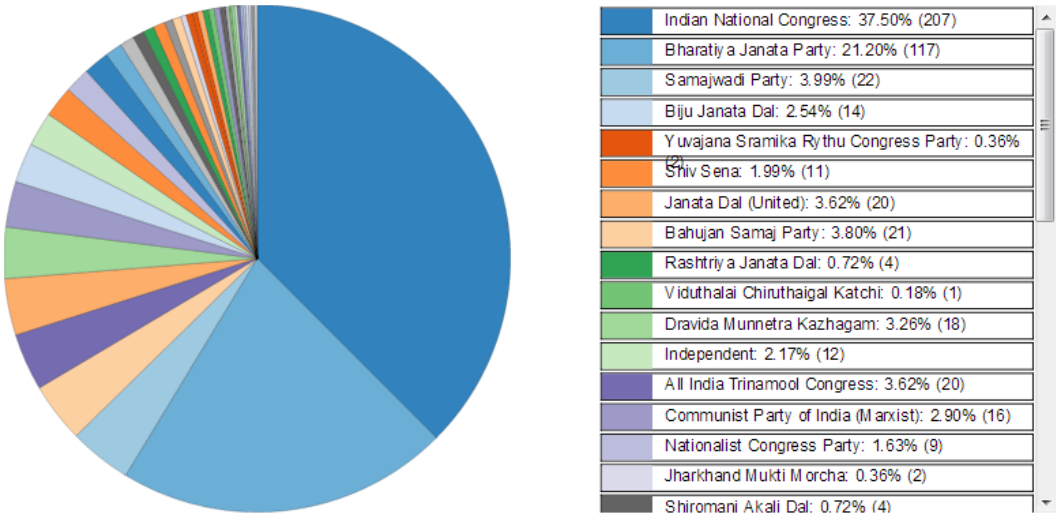

Fig 4.2.1: Political Parties in India

The region-wise distribution of political parties is as follows:

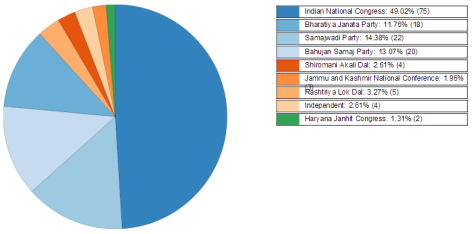

Fig 4.2.2: North India

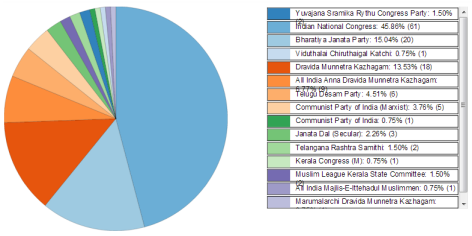

Fig 4.2.3: South India

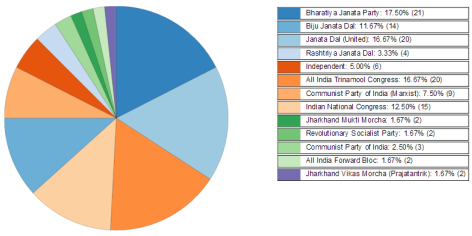

Fig 4.2.4: East India

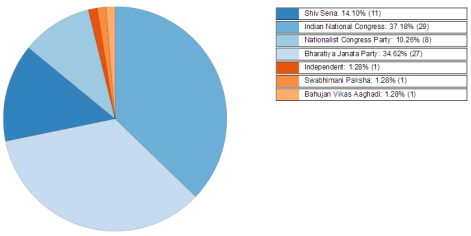

Fig 4.2.5: West India

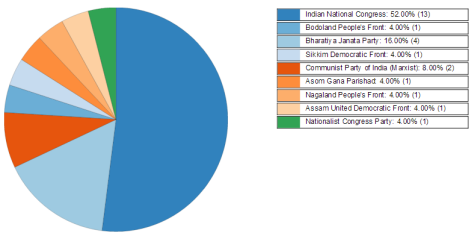

Fig 4.2.6: North East India

From the regional plots of political parties, it is clear that the Indian National Congress enjoys more than 50% patronage in North, South and North-East India. The BJP on the other hand draws its major support from the Western part of India where it has shown an incline towards developmental politics with Narendra Modi at the fore. In the East, regional parties like Trinamool Congress draw prominence.

Educational Qualifications

We see that contrary to popular perception, most Parliamentarians in India are well educated. There are 257 Graduates, 147 Post Graduates and 26 Doctrates in the parliament.

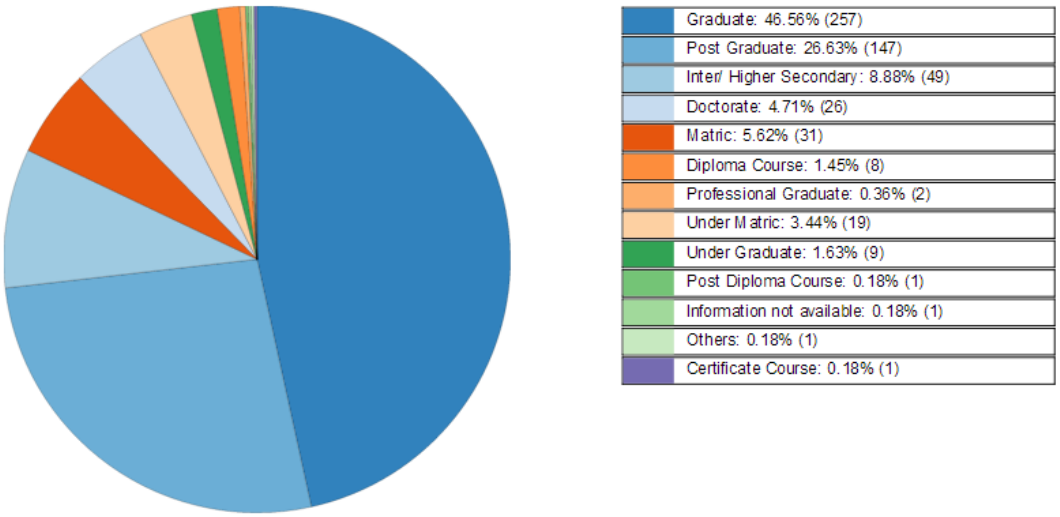

Fig 4.2.7: Educational Qualifications of Parliamentarians

Gender

The all India distribution of sex in the Parliament is as follows:

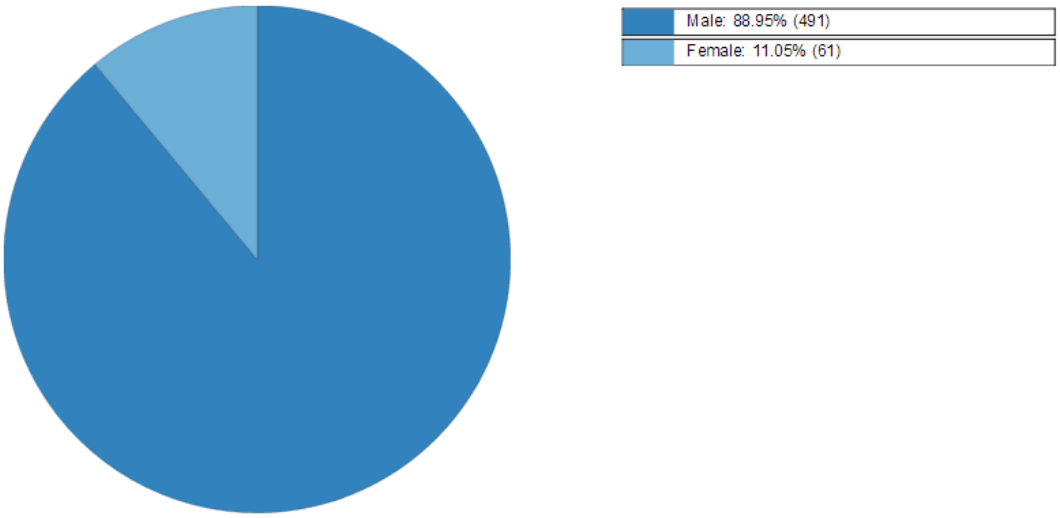

Fig 4.2.8: Gender in Indian Parliament

The regions of interest as far as Gender is concerned turn out to be North India and South India.

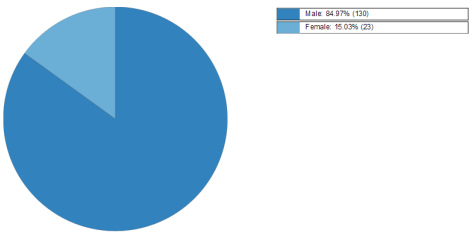

Fig 4.2.9: North India

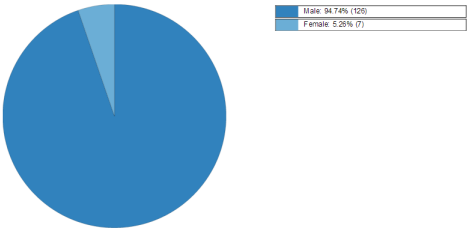

Fig 4.2.10: South India

We clearly see a prejudice against women in the South with only 5.26% of all seats going to the fairer sex. In the North however, the percentage of women stands at 15.03% - significantly higher than the national average. This gives us an idea as to the nature of the societies in the two regions and the amount of liberalization that women enjoy in the societies.

4.3 Area Chart

This chart shows the statewise distribution of political parties in a much more visually simulating manner.

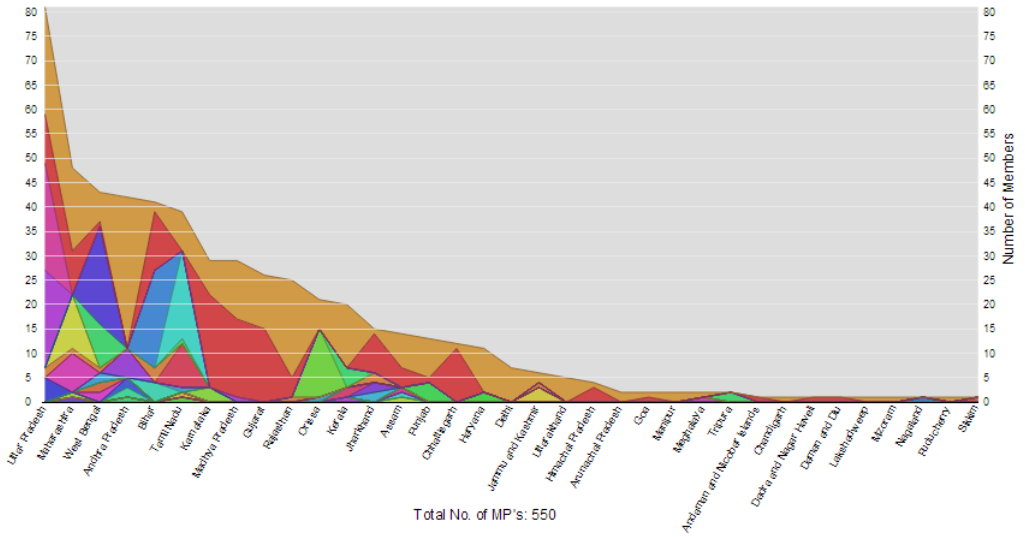

Fig 4.3.1: Distribution of Political Parties in States

It is also seen that there are mainly two parties that have a truly National character: BJP (shown in Red) and Indian National Congress (shown in Orange).

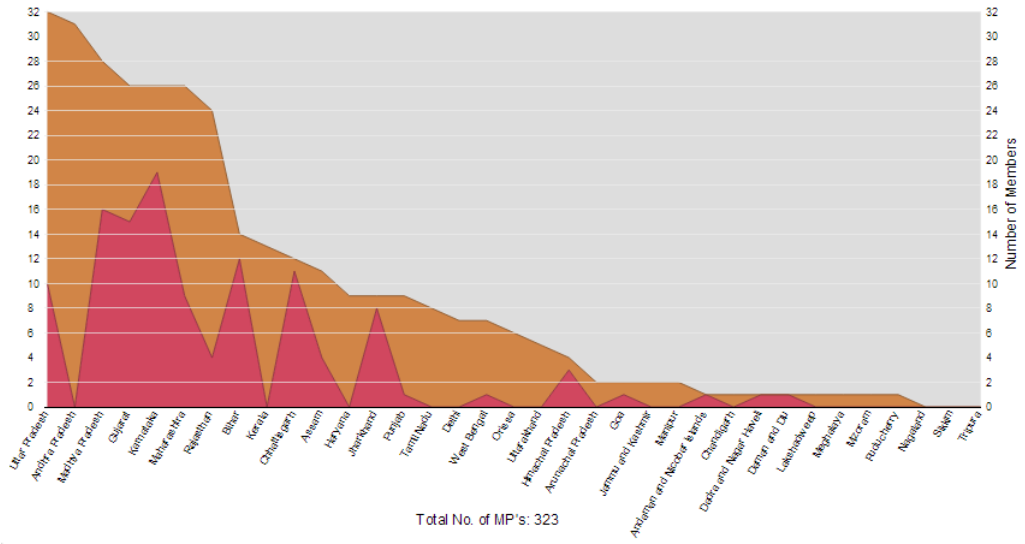

Fig 4.3.2: Distribution of Congress and BJP in States

4.4 Bar Chart

The main purpose of the bar Charts is to verify the truth of the conclusions above. However a few interesting insights into the nature if the Parliament were also obtained as is evident from the following:

Educational Qualifications

When we plot the Educational Qualifications of Members against the average age of the Parliamentarians with that qualification, we obtain the following plot:

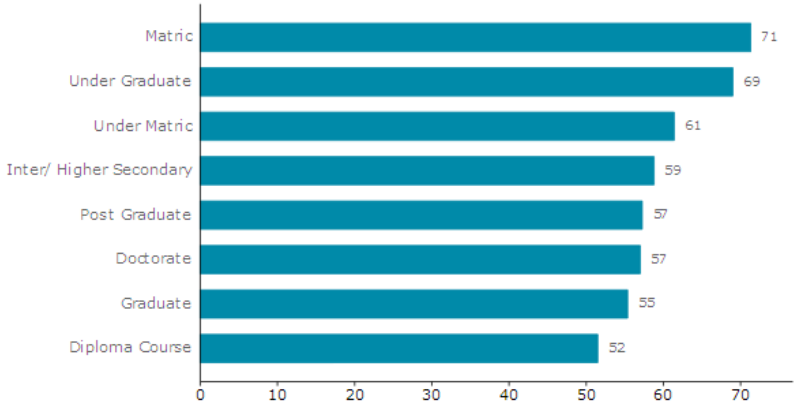

Fig 4.4.1: Educational Qualifications of MPs

From this graph, we can conclude that the popular perception of Indian Parliament being full of illiterate people used to be true in the past. But the young and dynamic new world leaders that are coming up are more educated and possess skills that could be very useful in policy making and problem solving.

4.5 Scatter Plot Matrix

We have created a 6-dimensional Scatter Plot between Age, Debates, Questions, Attendance, Educational Qualifications and Private Member Bills. Selection in any one graph

highlights the selection in all other graphs and allows us to draw some beautiful insights regarding the nature of the parliament. A screenshot of the Matrix is as shown below:

Fig 4.5.1: Scatter Plot Matrix

4.6 State-Party Mashup

A Mashup was designed between the states and political parties. In accordance with the nature of Indian Democracy, we see that the mesh is very sparsely populated with only the Congress and the BJP depicting a truly national character. The relevant cells of the mashup have been coloured in scale relative to each other in regard to the field selected. The sparse nature of the graph discourages any meaningful interpretations to be drawn from the graph. However, one can play around with the graph sorting it as per his wishes or choosing parameters on which to base the scaled colouring and draw his own conclusions regarding the nature of the Indian Parliament.

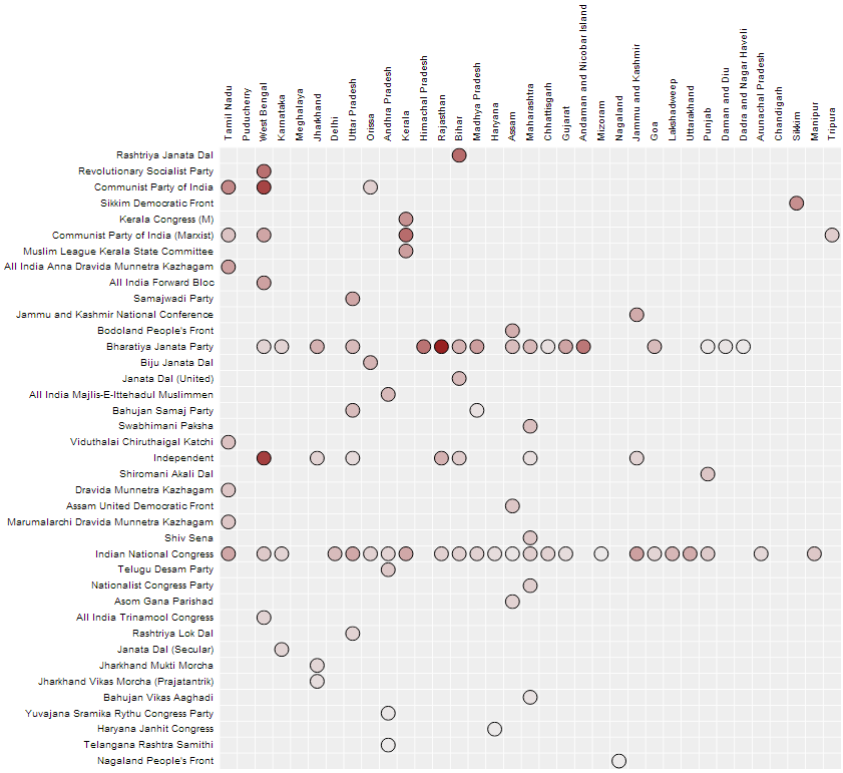

Fig 4.6.1: State-Party Mashup

4.7 Statewise Distribution

We have also created an intutive interface which allows the user to select whichever state of India he wants, and we will display information about how seats are distributed amongst the various political parties in that particular state.

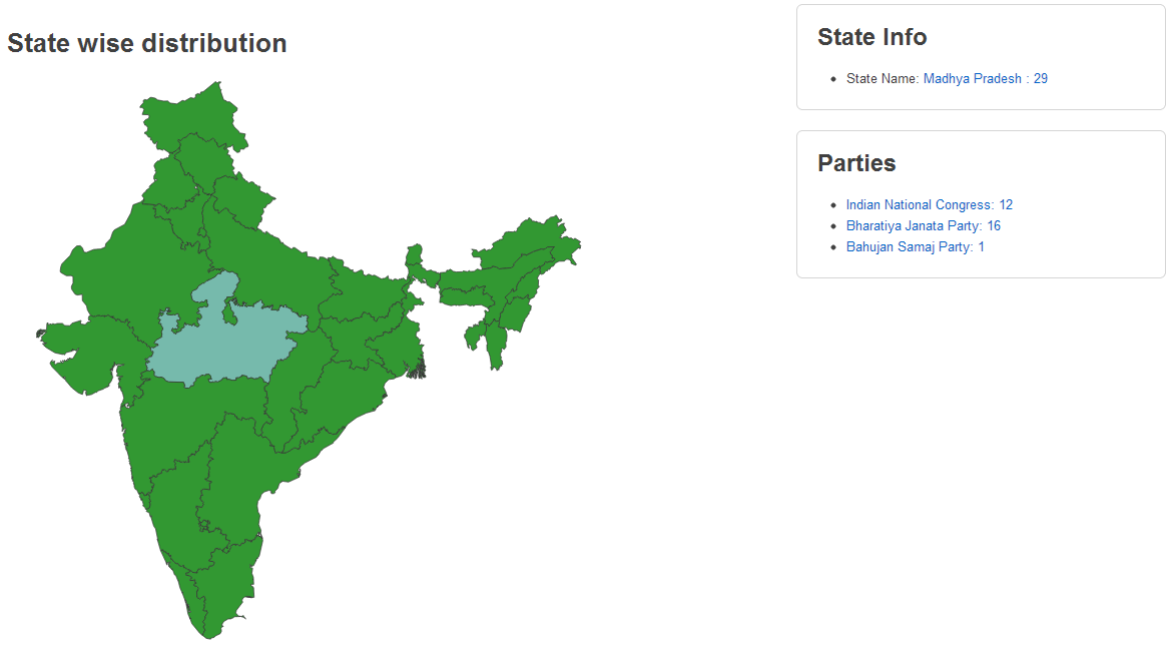

Fig 4.7.1: Statewise Distribution

5 Hypothesis Testing

Various performance parameters available to us for all the Members of the Parliament were - attendance, number of questions asked, number of private bills introduced and the number of debates participated in. We have done our analysis in the following major categories. We have tried to gauge the impact of various parameters on the performance of an MP. Classification has been done on the basis of:

5.1 Age

We divided the dataset into two parts - young and old. The barrier between the two classes was erected at the average age of all members, which was found to be 56.02 years. Thus, it was found that there are 256 young MPs and 239 old MPs that are not ministers in the Indian Parliament and for whom performance data is available. We did analysis on the following parameters:

Attendance

It was expected that younger members of the parliament would be more active and thus would have higher attendance. In order to test our hypothesis rigorously, we decided to do a t-test. The t-value obtained was 3.23 with a degree of freedom $\nu = 493$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong. However, what was shocking was that *it is not the youngsters that attended more but actually the older members.* The confidence level in this regard was 99.9%.

Debates

Since the policies being discussed in the Parliament would directly affect them to a larger extent, the younger members were expected to be more involved in the debating procedure. In order to test our hypothesis rigorously, we decided to do a t-test. The t-value obtained was 1.69 with a degree of freedom $\nu = 447$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong. However, what was shocking was that it is not the youngsters that debated more but actually the older members. Thus *elder members take part in more debates in the parliament than their younger counterparts.* The confidence level in this regard was 95%. Their experience and the will to change society must be driving them to take part in debates.

Private Member Bills

It was expected that with all their years of experience behind them, the elder members of the Lok Sabha would have introduced more Private Member Bills in the Parliament. In order to test our hypothesis rigorously, we decided to do a t-test. The t-value obtained was 1.39 with a degree of freedom $\nu = 420$. The t-value was less than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is correct i.e. *the number of private bills introduced by a member is independent of his age and experience.*

Questions

It was expected that younger members would be more active in the Parliament and thus would ask more questions. Also, since the policies being discussed in the parliament would

directly affect them to a larger extent, they were expected to be more involved. In order to test our hypothesis rigorously, we decided to do a t-test. The t-value obtained was 1.62 with a degree of freedom $\nu = 488$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *the number of questions asked by youngsters are more than those by the elder members.* The confidence level in this regard was 94.5%.

5.2 Educational Qualifications

The MPs were divided into two groups based on the degree to which they were educated. Those with educational qualification as Inter/Higher Secondary or below were classified to be less educated, while aall others were classified to be highly educated. Going by this classification, there are 95 less educated MPs and 397 highly educated members in the parliament. We did analysis on the following parameters:

Attendance

We were expecting that more educated members of the parliament would show more sincerity towards the Parliament. In order to rigorously test our hypothesis, we carried out a t-test. The t-value obtained was 1.68 with a degree of freedom $\nu = 128$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *the more educated a member is, the more likely is he to attend the sessions of the Parliament.* The confidence level in this regard was 95%.

Debates

It was expected that the more educated members of the parliament would take part in the deliberations of the parliament with greater frequency. In order to rigorously test our hypothesis, we carried out a t-test. The t-value obtained was 3.43 with a degree of freedom $\nu = 208$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *the more educated a member is, the more likely he is to participate in deliberations on the floor of the Parliament.* The confidence level in this regard was 99.95%.

Private Member Bills

It was expected that the more educated members of the parliament would introduce much greater number of Private Member bills in order to bring changes in the nature of the Indian Society. In order to rigorously test our hypothesis, we carried out a t-test. The t-value obtained was 0.35 with a degree of freedom $\nu = 111$. The t-value was less than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is valid i.e. *the educational qualifications of an MP does not affect the number of Private Bills introduced by him.*

Questions

It was expected that the more educated members of the parliament would stand in a better position to ask questions upon the various decisions of the Government. In order to rigorously test our hypothesis, we carried out a t-test. The t-value obtained was 1.16 with a degree of freedom $\nu = 143$. The t-value was less than the critical t-value at this value

of ν. Thus, we conclude that the null hypothesis is valid i.e. *the educational qualifications of an MP does not affect the number of questions he asks.*

5.3 Political Alliances

The nature of Indian Democracy is such that any analysis of the Parliament is incomplete without giving due consideration to the various alliances present in the Parliament. At present, there are mainly two alliances in the Parliament:

1. **United Progressive Alliance (UPA):** This is the political alliance that is currently in power. It consists of the following main political parties: Indian National Congress, Nationalist Congress Party, All India Majlis-E-Ittehadul Muslimmen, Assam United Democratic Front, Dravida Munnetra Kazhagam, Jammu and Kashmir National Conference, Kerala Congress (M), Muslim League Kerala State Committee, Rashtriya Lok Dal, Sikkim Democratic Front and Viduthalai Chiruthaigal Katchi.

2. **National Democratic Alliance (NDA):** This is the political alliance that forms the main opposition. It consistes of the following main political parties: Asom Gana Parishad, Bharatiya Janata Party, Haryana Janhit Congress, Janata Dal (United), Jharkhand Mukti Morcha, Nagaland People's Front, Shiromani Akali Dal, Shiv Sena and Telangana Rashtra Samithi.

In all there are 197 members of the UPA and 159 members of the NDA that are not ministers in the Lok Sabha. We did analysis on the following parameters:

Attendance

We wanted to find out whether there existed any relationship between the attendance of a party and whether it was in power. For this, we conducted a t-test between the two. The t-value obtained was 2.18 with a degree of freedom $\nu = 305$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *the parties in power maintain higher attendance.* The confidence level in this regard was 98.5%.

Questions

The parties in the opposition were expected to be asking more questions in order to show inconsistencies in the policies of the government. We conducted a t-test to confirm our hypothesis. The t-value obtained was 1.56 with a degree of freedom $\nu = 345$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *the parties in opposition ask more questions.* The confidence level in this regard was 94%.

Debates

The parties in the opposition were expected to be taking part in more debates in order to show inconsistencies in the policies of the government and show the supremacy of their own policies. We conducted a t-test to confirm our hypothesis. The t-value obtained was 2.31 with a degree of freedom $\nu = 282$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *the parties in opposition participate in more debates.* The confidence level in this regard was 99%.

Private Member Bills

The parties in power were expected to be tabling more Private Member Bills. Otherwise, either the government would not continue to stay in power or the members would not belong to different alliances. We conducted a t-test to confirm our hypothesis. The t-value obtained was 0.51 with a degree of freedom $\nu = 316$. The t-value was less than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is correct i.e. *the number of private member bills tabled by a member does not depend on whether he is in power or not.*

Educational Qualifications

We were expecting that with increase in voter awareness, the party that fields higher number of well educated candidates was more likely to come into power. To test our hypothesis, we conducted a t-test on the two opposing alliances and the degree of educational qualification of their elected members. The t-value obtained was 4.68 with a degree of freedom $\nu = 265$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *the greater the number of educated candidates fielded by an alliance, greater are the chances that it would come into power.* the confidence level in this regard was 99.999%. This is probably because the masses understand that the more educated their representative is, the more capable he is of resolving local issues they face on a daily basis.

Age

In the past few years, there has been increased clamour for the younger generation to play an important role in public life. So we decided to do a t-test in order to test the political fortunes of political parties vis-a-vis the age of their candidates. The t-value obtained was 0.46 with a degree of freedom $\nu = 342$. The t-value was less than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is correct i.e. *the age of its candidates does not affect the political fortunes of an alliance in the elections.* Thus , we can say that all the talk of younger politicians is a sham and that ground realities are very different from what is presented to us by the mass media.

5.4 Gender

This classification basically served to test the various *gender based stereotypes* prevalent in Indian society. There are in all 53 females and 441 males that are not ministers in the Lok Sabha (for whom performance data is available). We did analysis on the following parameters:

Attendance

In Indian society, females are considered to be more sincere in their work and tend to take their profession more seriously. Thus, going by the stereotypes, we were expecting that the attendance of females would be much greater than that of their male counterparts. In order to rigorously test this hypothesis (or stereotype as we may call it), we decided to do a t-test. the t-value obtained was 0.43 with a degree of freedom $\nu = 66$. The t-value was less than the critical t-value at this value of ν. Hence we can say that the null hypothesis is correct i.e. *attendance of members is not dependent on their gender.*

Educational Qualifications

In the Indian society, the common view is that males are more educated them females. In order to rigorously test this hypothesis, we decided to do a t-test. There are 60 females and 490 males for whom educational qualification data is available. The t-value obtained was 0.52 with a degree of freedom $\nu = 74$. The t-value was less than the critical t-value at this value of ν. Hence we can say that the null hypothesis is correct i.e. *Educational Qualifications of members do not depend on their gender*, atleast in the context of the Indian Parliament.

Debates

The Society generally paints the picture of females being more talkative in nature and interested in debates. We decided to gauge how much truth does this stereotype hold when talking of the Indian Parliament. So, we decided to do a t-test. the t-value obtained was 1.36 with a degree of freedom $\nu = 81$. The t-value was less than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is correct i.e. *Participation in debates by members is not dependent upon their gender* in the Indian Parliament.

Private Member Bills

The general viewpoint is that males are more proactive with the introduction of new policies. We wanted to test whether the same is reflected in the Indian Parliament. So, we decided to do a t-test. the t-value obtained was 1.66 with a degree of freedom $\nu = 139$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong i.e. *males introduce more private bills in the parliament than their female counterparts*. The confidence level in this regard was 95%.

Questions

The society associates a quality of inquisitiveness with women. We wanted to find out whether this alleged nature of theirs holds true when extended to the Indian Parliament. So, we decided to do a t-test. The t-value obtained was 1.82 with a degree of freedom $\nu = 70$. The t-value was more than the critical t-value at this value of ν. Thus, we conclude that the null hypothesis is wrong. However, what was shocking was that it is not the females that asked more questions but actually the males. Thus *males ask more questions in the parliament than their female counterparts*. The confidence level in this regard was 96%.

5.5 Geographical Boundaries

India is the seventh-largest country in the world with a wide variety of Geographical features and regions. We have done our analysis on the attendance in the Parliament based on the following parameters:

Region

We considered the two geographically distinct regions of North and South India separately and carried out analysis on them. The states which were considered to be a part of North India were: Delhi, Haryana, Chandigarh, Himachal Pradesh, Jammu & Kashmir, Madhya Pradesh, Rajasthan and Uttar Pradesh. On the other hand, constituents of South India were taken to be: Andaman & Nicobar Islands, Andhra Pradesh, Karnataka, Kerela,

Lakshdweep and Tamil Nadu. There are in all 162 members from North India and 116 members from South India who are not Ministers.
We had expected that North Indian MPs would have a higher attendance than the South Indians owing to the large distances between the Parliament House in Delhi and South India in addition to the general non-favorable climes of the North. We conducted a t-test to check our hypothesis. The t-value obtained was 4.38 with a degree of freedom $\nu = 222$. The t-value was more than the critical value at this value of ν. Thus, we conclude that the null hypothesis is incorrect i.e. *MPs from North India attend the Parliament more than those from South India.* The confidence level in this regard was 99.999%.

State Size

We found the average number of MPs elected per state (calculated to be 15.71). All those states that had a higher representation than this average value (12 of them) were classified to be large states, while the rest of them were classified as small states. It was found that 400 members belonged to large states and 92 members belonged to the small states and were also not ministers in the government. We expected that small states would in general have higher attendance so that their viewpoint is always considered in the parliament. To check our hypothesis, we conducted a t-test. The t-value obtained was 0.61 with a degree of freedom $\nu = 118$. The t-value was less than the critical value at this value of ν. Thus, we conclude that the null hypothesis is correct i.e. *the size of a state does not affect its attendance in the parliament.*

6 Conclusions

We can conclude after a rigorous analysis of all fronts of the Parliament of India that the Parliament provides a lot of interesting insights about the Indian society as a whole. It proves many of our stereotypes of the society at large to be correct while vehemently discarding a few of them. We tested a large number of hypotheses and subjected them to the Welch's t-test.

All of this helped us to gain an insight into the nature of Indian Democracy and thus the Indian society as a whole. Much research can still be done in this area by taking into consideration the nature of the Lok Sabha over the years from 1947 onwards. However, that lies beyond the ambit of this project.

• • •

Report developed in LaTeX