Celia Heudebourg
Minju Kim
Corey McGinnis
MATH 155: Final Project

Distorting Democracy:
How Gerrymandering Skews the Composition of the House of Representatives

Introduction

Do you think your vote mattered the last time you cast a ballot? The answer might have more to do with geography than you would expect. It's possible that the congressional district you voted in was already set to turn red or blue before you and every other person voting there got to the polling place. This project will seek to analyze, quantify, and illustrate the role that gerrymandering has played in determining voting patterns by answering the following research question: To what extent has the compactness of congressional districts since 2010 affected the relationship between votes cast in the 2012 and 2014 elections and the makeup of the members of the U.S. House of Representatives?

We chose this research question with the hope that it would provide a comprehensive analysis of how gerrymandering has affected political representation. Before collecting the data, we expected our analysis to show a very strong linear relationship. We thought the effects of gerrymandering would be more pronounced. Our research has led us to the conclusion that there is a linear relationship between the compactness of a district and absolute difference of votes cast for 2012, but there is not for 2014. In other words, we reject the null hypothesis for 2012 but fail to do so for 2014. These results can be generalized to the U.S. population as a whole, except for states with homogeneous representation which are not included in this study. The terms *absolute difference* and *homogenous representation* will be defined further in the paper.

Background

The United States has three branches of government; the executive branch, the judicial branch and the legislative branch. The legislative branch consists of two chambers; the Senate, and the House of Representatives. The Senate is made up of two senators from each state, while each state's representation in the House is proportional to its population. The states themselves are broken up into congressional districts similar in population, with one representative for each district. This way, each representative in the House is accountable for their constituents in one small region. Because of this set up, the House of Representatives is arguably the most democratic body in the United States government.

From a young age, people in the United States are taught that they live in a democracy. On the surface this is true. All citizens eighteen years or older have the right to vote (with some exceptions) and voters determine who will represent them in elected office. However, the political mechanisms that make this democracy work are often less fair and balanced than they appear to be. Gerrymandering, a practice that establishes a political advantage for a particular party by manipulating district boundaries, has become all too common in recent elections.

Before a single vote is cast, cartography, not voters, can be the deciding factor in whether a district goes to the Democrats or the Republicans. One example of how legislatures sway these votes is by drawing district borders around demographic minorities (or other groups of people known to vote in a certain way) in order to both decrease and increase minority representation in state governments and congressional delegations.

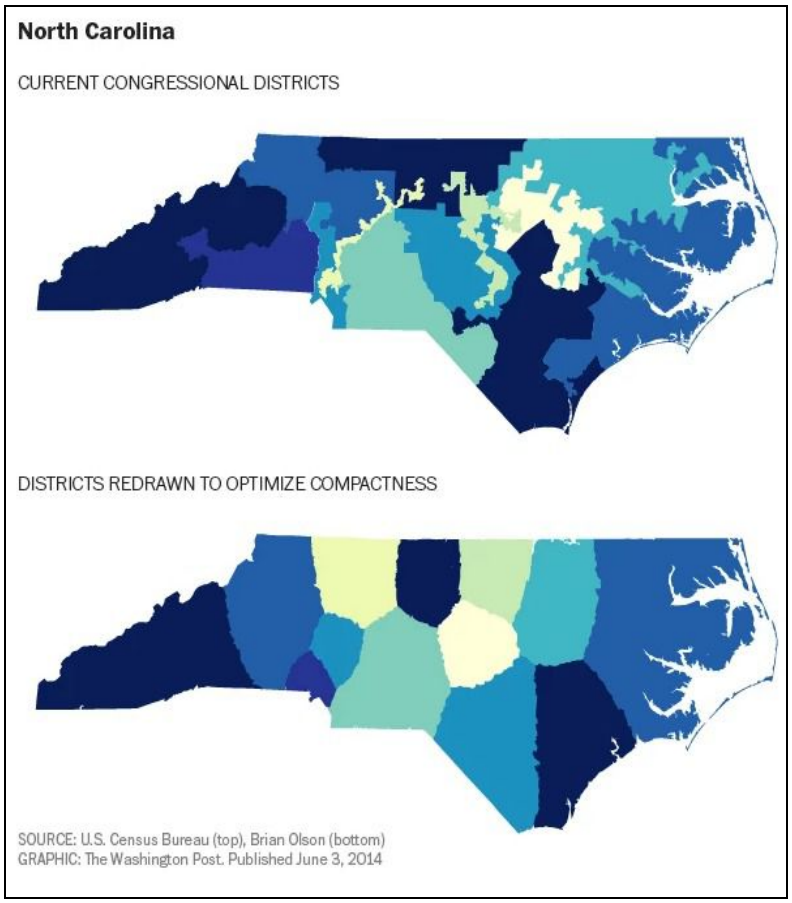


Fig 1. - A map of North Carolina's current congressional districts and a map showing what North Carolina's districts would look like if they were redrawn to maximize compactness.
https://newsela.com/articles/gerrymandering-redistricting/id/14309/

Gerrymandering, as a politically-biased practice, is not highly monitored, even in those states that set compactness requirements. The degree to which a district has been gerrymandered can be calculated with one of several measures. Four of the most common measures of compactness are the Roeck, Convex Hull, Polsby-Popper, and Schwartzberg measures. These measures calculate the compactness of a district in different ways, resulting in different accounts of how spread out or closely-drawn a district is. A narrow and drawn-out, gerrymandered district could score a high compactness score because of other topographical or demographic factors, thereby underestimating how gerrymandered that district is.

For the purposes of this research, we also define two new terms. A state with *homogeneous representation* is a state whose representatives all come from one party. *Absolute difference* is the absolute value of the difference between the percentage of votes cast for a party, and the percentage of seats in the House of Representatives occupied by that party. Because the United States has a two-party system, this value is the same whether you calculate this for the Democratic party or the Republican party. We are looking at absolute differences on a state-by-state basis. For example, suppose State A has 10 representatives. If seven of these representatives are members of the Democratic Party, 70 percent of the seats for State A are occupied by Democrats. In the most recent election, 60 percent of voters voted for Democratic candidates. In this case, our absolute difference would be 0.10, which is the absolute value of 0.60 minus 0.70.

Methods

Due to the national scope of our research, we opted to collect our data from a number of online sources. The House election results in 2012 were found in the Federal Election Commission's official website (link: http:// www.fec.gov/pubrec/electionresults.shtml). The House election results in 2014 were found in a New York Times article which got the result from the Associated Press (link: http://elections.nytimes.com/2014/results/house). We found whether or not states had compactness requirements on a university website entitled "All about redistricting" (link: http://redistricting.lls.edu/where-tablefed.php). The different measures of compactness were found on two websites: "Redistricting the Nation" (link: http://www.redistricting thenation.com/) and "Governing: the States and Localities" (link:http://www.governing.com/gov-data/politics/gerrymandered-congressional-districts-compactness-by-state.html).

Based on these data, we were able to calculate the other variables that we needed for this project. There are both categorical and quantitative variables included in the research. Categorical variables include: whether the state has a compactness requirements (yes/no), and the political homogeneity of the state (yes/no). Quantitative variables include: the number of votes cast in the 2012 election by party (unit: vote), the number of votes cast in the 2014 election by party (unit: vote), the absolute differences for 2012 and 2014, and the four measurements of compactness (PolsbyPopper, Shwartzberg, ConvexHull, Reock). Our sample size was initially every state in the country, however, we removed the states that had homogenous representation. This decision regarding the sample size was made solely with the intention of targeting our data to the purposes of our study. States with homogenous representation or a single district cannot be gerrymandered to affect votes.

Because we are using government and university-produced data, we as researchers have not introduced personal biases during the collection of the data. The only potential for sampling bias would be because we selected the websites from which we collected the data. However, since we used government data, we find our sources to be trustworthy and objective. This project was not funded. We received zero compensation for this research.

- **Methods of statistical analysis**

We used the 0.99.491 version of RStudio. The research primarily utilized the linear regression model in order to determine the linear relationship between the *absolute difference*, and the compactness of a state. We used a significance level of 0.2, so we have a 20 percent probability of falsely rejecting the null hypothesis. We chose this significance level because for this research, a Type II error, in which we fail to reject the null hypothesis, is worse than a Type I error, in which we falsely reject the null hypothesis.

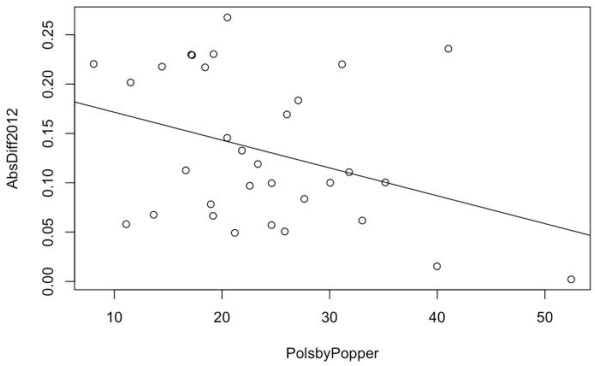
To model and visualize our data, we used "lm" and "anova."

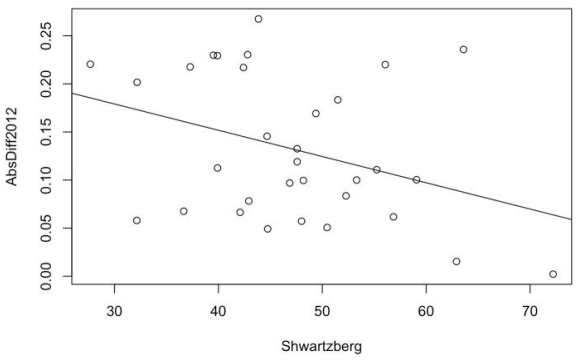
Results

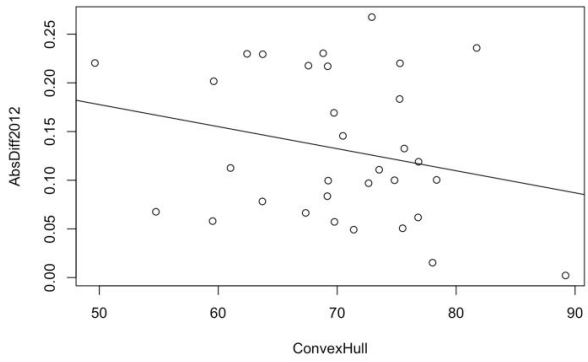
- **Plots**

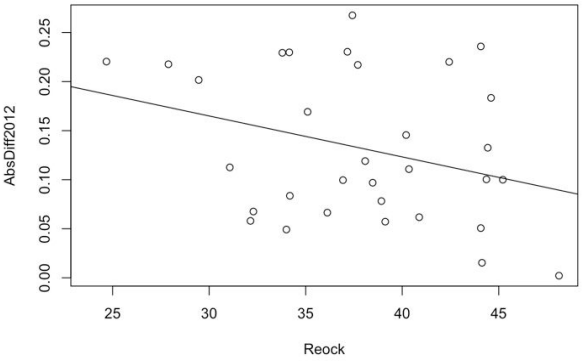
The plots below explore the linear relationships between the absolute differences for each state in our dataset, and each state's compactness based of the five different measures we have selected. We depict these relationships for both the 2012, and the 2014 elections.

2012 House Election Results

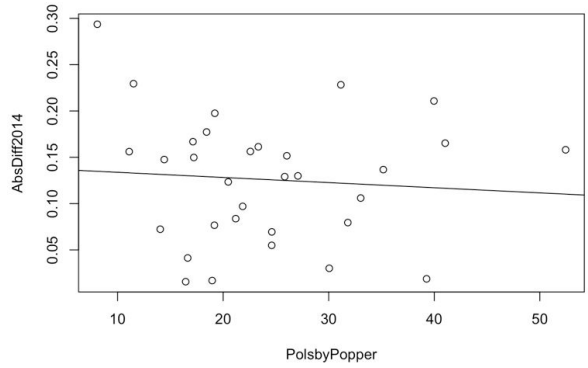


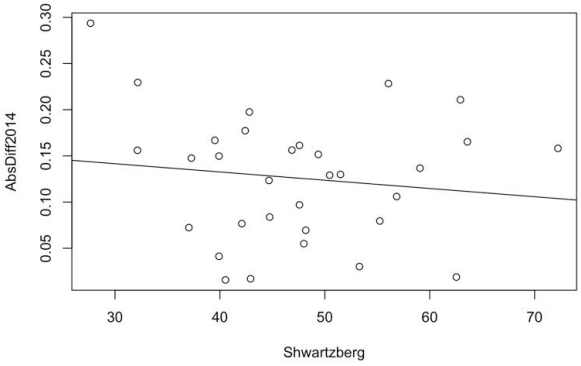


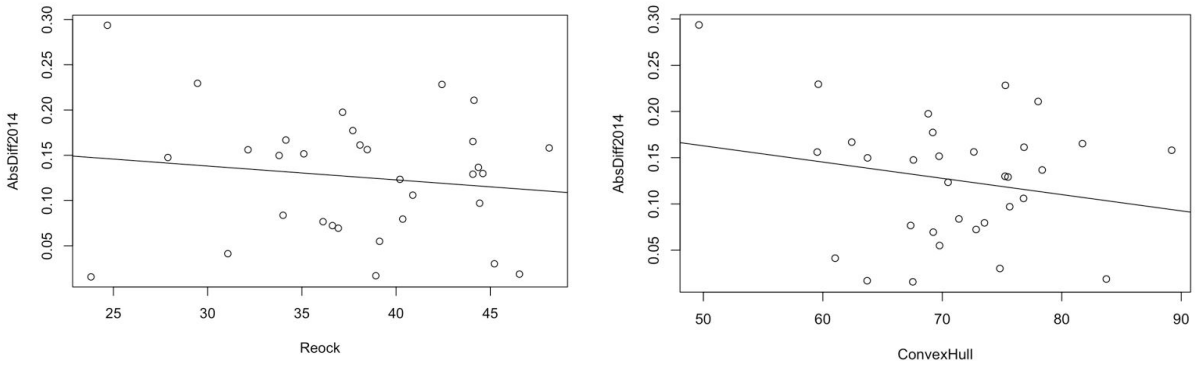




2014 House Election Result







In all eight plots, the data points were somewhat scattered, but each displayed a downward trend, indicating that as the congressional districts become more compact, we expect the absolute difference to decrease. This downward slope is greater in the plots for our 2012 data. Our axes do not contain units because the data for our variables (compactness coefficients and the difference between percentages) have no units. These plots suggest that gerrymandering does have negative impact on the fairness of the results of our elections.

- **Summary statistics and beginner exploratory data analysis**

The research is based on election information from 38 states, where states with homogeneous representation are excluded (the exclusion is explained in the discussion page). Our two response variables, the absolute difference in 2012, and the absolute difference in 2014 represent the absolute difference in a state in each election year. Therefore, the linear relationship between the absolute differences and the corresponding measures of compactness illustrates the association between compactness and the accuracy of representation by political party.

In 2012, the mean absolute difference was 0.129443. This means that the mean difference in proportion of votes and representation was 12.9 percent. We found a negative regression line for all four measures, which means that as compactness increases, absolute difference decreases. Values for the adjusted r-squared are 0.09781, 0.09527, 0.02709, 0.06648 for PolsbyPopper, Shwartzberg, ConvexHull, and Reock, respectively. This shows that two to nine percent of the response variable can be explained by the explanatory variable.

In 2014, the mean absolute difference was 0.12342. We found a negative regression line in all four measures, which means that as compactness increases, absolute difference decreases. Values for the adjusted r-squared are -0.02651, -0.01512, 0.009642, -0.01296 for PolsbyPopper, Shwartzberg, ConvexHull, and Reock, respectively. This shows that one to two percent of the response variable can be explained by the explanatory variable. Overall, even though we recognize the negative relationship between the two variables, the statistical significance of the linear relationship is dubious. We discuss this issue at greater length in the hypothesis testing.

- **Model Construction**

These models measure the relationship between the aggregate compactness of the congressional districts in a state, and the absolute difference. We had eight different models: four for each election year (2012 and 2014), each using a different measure of compactness

(PolsbyPopper, Shwartzberg, ConvexHull, and Roeck). There is no interpretation for any of the intercepts because no state can have an aggregate compactness of zero. All eight models are similar in that they predict that the absolute difference will decrease as compactness increases. This suggests that states with less gerrymandering will have more democratic representation in the House of Representatives.

PolsbyPopper 2012

$AbsDiff2012 = 0.199725 - 0.002823 * PolsbyPopper$

An increase in compactness by 1 Polsby-Popper unit predicts a decrease in the absolute difference by 0.002823 units in the 2012 Congressional Elections. Our multiple R-squared is 0.1269, so this model predicts 12.69% of the variability.

Shwartzberg 2012

$AbsDiff2012 = 0.261028 - 0.002731 * Schwartzberg$

An increase in compactness by 1 Schwartzberg unit predicts a decrease in the absolute difference by 0.002731 units in the 2012 Congressional Elections. Our multiple R-squared is 0.1245, so this model predicts 12.45% of the variability.

Convex Hull 2012

$AbsDiff2012 = .291173 - 0.002268 * ConvexHull$

An increase in compactness by 1 Convex Hull unit predicts a decrease in the absolute difference by 0.002268 units in the 2012 Congressional Elections. Our multiple R-squared is 0.05847, so this model predicts 5.85% of the variability.

Roeck 2012

$AbsDiff2012 = .290195 - 0.004174 * Roeck$

An increase in compactness by 1 Roeck unit predicts a decrease in the absolute difference by 0.004174 units in the 2012 Congressional Elections. Our multiple R-squared is 0.09659, so this model predicts 9.66% of the variability.

PolsbyPopper 2014

$AbsDiff2014 = 0.1394739 - 0.0005575 * PolsbyPopper$

An increase in compactness by 1 Polsby Popper unit predicts a decrease in the absolute difference by 0.0005575 units in the 2014 Congressional Elections. Our multiple R-squared is 0.006605, so this model predicts 0.661% of the variability.

Shwartzberg 2014

$AbsDiff2014 = 0.1684106 - 0.0008947 * Schwartzberg$

An increase in compactness by 1 Schwartzberg unit predicts a decrease in the absolute difference by 0.0008947 units in the 2014 Congressional Elections. Our multiple R-squared is 0.01763, so this model predicts 1.76% of the variability.

Convex Hull 2014

$AbsDiff2014 = 0.250894 - 0.001760 * ConvexHull$

An increase in compactness by 1 Convex Hull unit predicts a decrease in the absolute difference by 0.001760 units in the 2014 Congressional Elections. Our multiple R-squared is 0.04159, so this model predicts 4.16% of the variability.

Roeck 2014

$AbsDiff2014 = 0.184249 - 0.001535 * Roeck$

An increase in compactness by 1 Roeck unit predicts a decrease in the absolute difference by 0.001535 units in the 2014 Congressional Elections. Our multiple R-squared is 0.01972, so this model predicts 1.97% of the variability.

- **Hypothesis Testing**

Ho *(Null Hypothesis)*: There is no linear relationship between the aggregate compactness of a state's congressional districts and absolute difference.
HA *(Alternative Hypothesis)*: There is a linear relationship between the aggregate compactness of a state's congressional districts and absolute difference.
Significance Level: 0.2. We chose this significance level because for this research, a Type II error, in which we fail to reject the null hypothesis, is worse than a Type I error, in which we falsely reject the null hypothesis. Falsely rejecting the null hypothesis may lead to unnecessary funds being spent on independent redistricting committees. Failing to reject the null hypothesis could make the public indifferent to gerrymandering, in which case the disenfranchisement of voters would ensue.

We must assume that there is no linear relationship between compactness and absolute difference unless we find that the p-values for our models are below 0.2.

Models for the 2012 election

	PolsbyPopper	Shwartzberg	ConvexHull	Reock
p-value	0.0454	0.04766	0.1824	0.08339

Models for the 2014 election

	PolsbyPopper	Shwartzberg	ConvexHull	Reock
p-value	0.658347	0.46883	0.263	0.4433

We interpret each p-value as the probability that the variance of the data from our linear models is due to pure chance. Comparing the p-values of the measures of compactness, we rejected the null hypothesis for all four models for the 2012 election because their p-values were all less than 0.2. This means we can conclude that in 2012, the relationship between compactness and the

absolute difference is statistically significant. For the 2014 election, we failed to reject the null hypothesis for all four models because the p-values were all greater than 0.2. In other words, we conclude that there is no significant linear relationship between the absolute difference and the measurements of compactness for 2014.

Discussion

Our research found that in 2012, there was a statistically significant relationship between the aggregate compactness of a state's congressional districts and absolute difference, which is the absolute value of the difference between the percentage of votes cast for a party, and the percentage of seats in the House of Representatives occupied by that party. However, no statistically significant relationship was found for 2014. The results of this study can be generalized to the U.S. population as a whole, except for those states with homogeneous representation. While these results are somewhat contradictory, there may be an explanation for this difference between the data from 2012 and 2014. In recent electoral history, Democratic candidates have fared better in presidential election years (i.e. 2008, 2012), while Republican candidates had done better in midterm election years (i.e. 2010, 2014). Congressional districts were last redrawn after 2010, and because Republicans were successful in 2010, they got to draw the districts in many states. Therefore, it is possible that Republicans were able to gerrymander the districts in more states than Democrats. However, due to decisive Republican electoral victories in 2014, gerrymandering may have had less of an aggregate impact. Otherwise, it is possible that we made a Type II error and that for the 2014 data, there really was a linear relationship between compactness and absolute difference.

Looking at the result for the 2012 elections, it is possible to see just how the U.S. House of Representatives is not perfectly democratic. A linear relationship between compactness and absolute difference reveals that states with little to no gerrymandering have more democratic elections, and give more power to their voters. The mean difference between votes and representation was about 0.12 in 2014 and 0.13 in 2012, which shows that there is a difference between who we vote for, and who ends up in office. Clearly, some votes matter more than others. These results violate a democratic principle we hold dear: "one person, one vote." Given the linear relationship we found, a good way to alleviate this problem would be to impose compactness requirements in all states, to reduce the possibility of gerrymandering and voter disenfranchisement.

It is also important to note that there are a number of other factors besides the compactness of congressional districts that can distort representation in the House of Representatives. For example, liberal voters who tend to vote for Democratic candidates often pack themselves into urban areas, and are more likely to end up in a district with a high share of Democratic voters. Our data also revealed that the non-proportional allocation of votes exacerbates this problem of high absolute differences the most. Not surprisingly, the states with homogeneous representation had the highest absolute differences. We came to the conclusion that in states with homogenous representation, gerrymandering is not an issue. Because gerrymandering attempts to pack most of a party's supporters into as few districts as possible, its effect is that the party wins one district decisively, and loses the races in all other districts. This would not occur in a state with homogeneous representation, so including these states in

our models skewed our data. We decided it was appropriate to remove the data from these states.

It is also possible that the research had flaws in its design. There is a chance that for our 2012 data, we are making a Type I error, and there isn't actually a linear relationship between our variables. Furthermore, perhaps the absolute difference is not the best response variable to measure the fairness of elections. There may be other parameters that need to be taken into account. Future research could examine specific districts and look at how gerrymandering might impact and disenfranchise communities based on race, age, income, gender identity, national origin, or other factors. Our research was broad, so there may be a lot to learn from looking at specific cases and breaking them down into more categorical variables.

Appendix

All annotated R code used in the paper:

First of all, we attached the data into our file.

```
## attach(testData)
```

We created a grand mean model for absolute difference in 2012.

```
## grandmean <- lm(AbsDiff2012 ~ 1)
```

We created a linear regression model for absolute difference of 2012 and the four measures of compactness.

```
## ppmod12 <- lm(AbsDiff2012 ~ PolsbyPopper, data = testData)
## sbmod12 <- lm(AbsDiff2012 ~ Shwartzberg, data = testData)
## chmod12 <- lm(AbsDiff2012 ~ ConvexHull, data = testData)
## romod12 <- lm(AbsDiff2012 ~ Reock, data = testData)
```

We plotted the linear model into the graph.

```
## plot(PolsbyPopper, AbsDiff2012)
## plot(Shwartzberg, AbsDiff2012)
## plot(ConvexHull, AbsDiff2012)
## plot(Reock, AbsDiff2012)
```

We used summary command to get the coefficients and p-value of the models.

```
## summary(ppmod12)
## summary(sbmod12)
## summary(chmod12)
## summary(romod12)
```

We fit the regression line into the plots of individual measures.

```
## abline(0.199726, -0.002823)          >> For Polsby-Popper
## abline(0.2610277, -0.0027311)        >> For Shwartzberg
## abline(0.291173, -0.002268)          >> For Convex Hull
## abline(0.290195, -0.004174)          >> For Reock
```

We created a grand mean model for absolute difference in 2014.

```
## grandmean <- lm(AbsDiff2014 ~ 1)
```

We created a linear regression model for absolute difference of 2012 and the four measures of compactness.

```
## ppmod14 <- lm(AbsDiff2014 ~ PolsbyPopper, data = testData)
## sbmod14 <- lm(AbsDiff2014 ~ Shwartzberg, data = testData)
## chmod14 <- lm(AbsDiff2014 ~ ConvexHull, data = testData)
## romod14 <- lm(AbsDiff2014 ~ Reock, data = testData)
```

We plotted the linear model into the graph.

```
## plot(PolsbyPopper, AbsDiff2014)
## plot(Shwartzberg, AbsDiff2014)
## plot(ConvexHull, AbsDiff2014)
## plot(Reock, AbsDiff2014)
```

We used summary command to get the coefficients and p-value of the models.

```
## summary(ppmod14)
## summary(sbmod14)
## summary(chmod14)
## summary(romod14)
```

We fit the regression line into the plots of individual measures.

```
## abline(0.1394739, -0.0005575)        >> For Polsby-Popper
## abline(0.1684106, -0.0008947)        >> For Shwartzberg
## abline(0.250894, -0.001760)          >> For Convex Hull
## abline(0.184249, -0.001535)          >> For Reock
```