Project Documentation

Improve Visualization of Popular Support for Executive Actions in the U.S.

Yu Cai, Çağlar Otlu, Jens Rauenbusch

Introduction to Data Science – MIBMA 2020

Prof. Dr. Birkenkrahe

June 28, 2021

Table of Contents

Introduction **3**

Related Literature **4**

COVID-19 Relief Bill 4

Data Visualization 4

Research Question 5

Methodology **6**

Obtain and understand the data 6

Critique the existing data visualization 7

Rebuild the visualization 8

$1.9trn covid-19 relief (2021) 9

GOP Health Care Bill (2017) 18

Affordable Care Act (2009) 20

Bush Tax Cuts (2001) 23

Clinton Health Care Plan (1993) 30

Generic Function to display dotplot based on row numbers 36

Generic Function to display dotplot based on row numbers, choose y-value and add title 36

Visualising the averages, cleaning the data and graph 37

The main ggplot code 38

Plotting one bill with plot() and points() 39

Results **41**

Plotting all bills with ggplot2 41

Plotting all bills with plot() and points() 45

Discussion **47**

Limitations **48**

Conclusion **49**

References **50**

Appendices **51**

Appendix 1 51

Appendix 2 51

Introduction

The COVID-19 pandemic triggered a large number of government responses such as imposing public health measures, supporting high-speed development of vaccines, and providing massive financial aid for citizens, businesses, and public institutions. Many countries spent record amounts trying to balance out some of the effects of the pandemic, as did the United States (U.S.) under Democratic president Joe Biden.

In an article published in February 2021, The Economist proclaimed that “Joe Biden’s $1.9trn stimulus package is one of the most popular bills in decades” (The Economist, 2021). The article concludes that “with 70% of the public behind him, Mr Biden may not have to listen [to criticism by Republicans]” (ibid.).

Since this article is part of the *graphic detail* series[1], it contains a data visualization to support the argument. The original visualization is shown in Figure 1. It displays net support for 17 executive actions or bills from 1990 to 2021, taken from various polls on the bills.

Upon studying the visualization, we, the project team, were able to confirm the overall argument that the COVID-19 relief bill is one of the most popular ones. However, we could not verify the claim that 70% of the population support the bill. The graph has some additional shortcomings, such as a missing key for the light blue dots.

Overall, we believe that this piece of data journalism can be improved. Therefore, we chose to take a closer look at the data and their visualization as part of our exploratory data analysis (EDA) project.

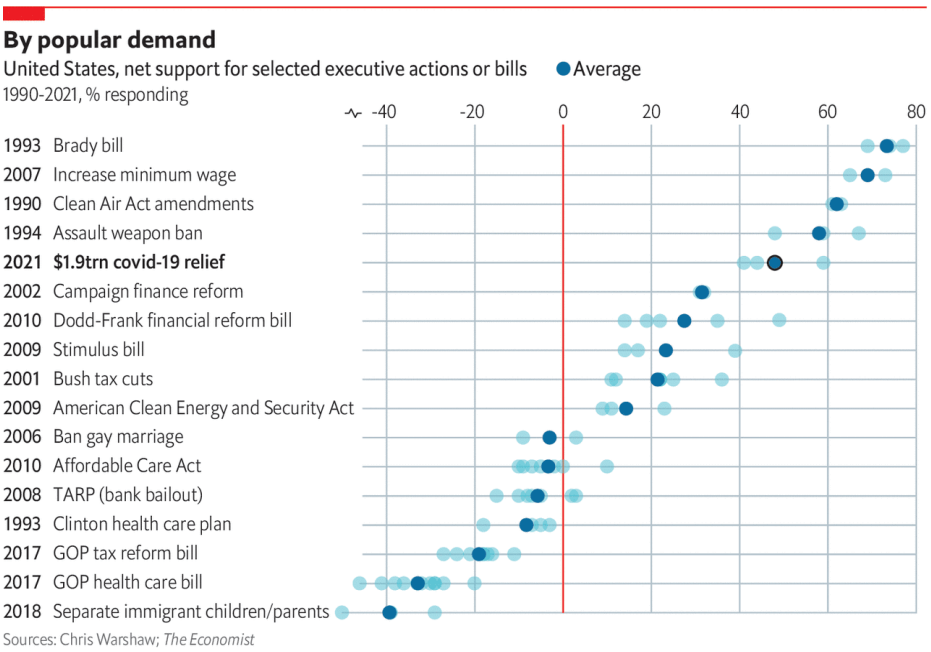


Figure 1. Original data visualization on net support for selected executive actions or bills in the U.S.
Source: The Economist (2021)

[1] *Graphic detail* is a series by The Economist which includes data journalism, infographics, and interactive information visualizations: https://www.economist.com/graphic-detail

Related Literature

COVID-19 Relief Bill

Biden's COVID-19 stimulus package, titled American Rescue Plan Act of 2021, is described as "one of the largest economic stimulus measures in American history, a sweeping $1.9 trillion COVID-19 relief bill" (Cornwell & Brice, 2021) (also see Figure 2). The measure included direct payments of $1,400 each to most U.S. citizens, $350 billion aid to governments, temporary relaxation of certain tax credits, and support for speedy distribution of vaccines (ibid.).

While the bill was positioned by most Democrats to be able to jumpstart the U.S. economy, many Republicans believed it too expensive and expected the economy to rebound without financial aid (ibid.). Therefore, the bill was heavily discussed in Congress and finally "came with zero Republican support after weeks of partisan debate and wrangling" (ibid.).

Figure 2. Tweet by president Biden, advertising the COVID-19 relief bill. Source: President Biden (@POTUS), 2021)

The public, however, generally favored the American Rescue Plan. In a survey by Data for Progress, a group of pollsters, the bill was found to be largely supported by likely voters across all parties (Penumaka, 2021). A national poll by Reuters/Ipsos also found that a majority of Americans were in favor of the bill (Cornwell & Brice, 2021).

Data Visualization

Data visualization is a process that builds on qualitative or quantitative data and results in their visual representation which can be read by an audience and allows communication or exploration of the data (Azzam et al., 2013).

This process is depicted by Carpendale (2003) as a series of iterations along five general steps (data or information, data representation, visual representation, visual presentation, view) which are connected through four transformative actions (data abstraction, visual abstraction, presentation transformation, interactive transformation) (see Figure 3).

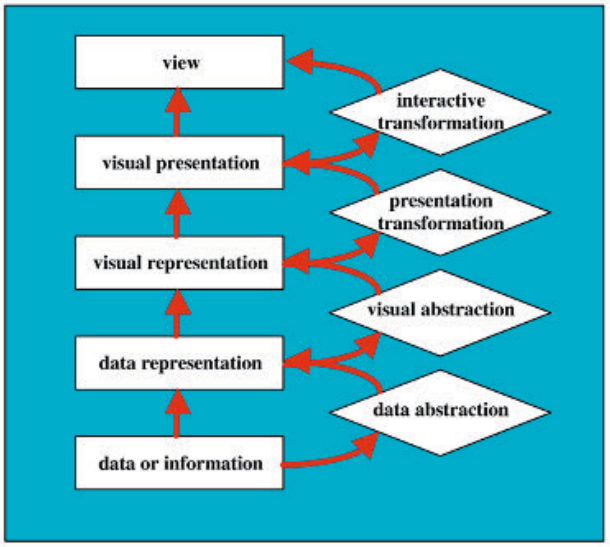


Figure 3. Visualization pipeline. Source: Carpendale (2003)

Dörk et al. (2013) introduce several principles for information visualization. While these are aimed at interactive visualizations, they nevertheless contain hallmarks against which any visualization may be evaluated:

- **Disclosure**: making decisions along the data visualization process transparent in order to create trust between authors and audience
- **Plurality**: allowing for a multitude of interpretations rather than limiting the visualization to one facet or view on a subject matter
- **Contingency**: enabling the audience to interact with and experience the visualization in order to foster understanding of the data and subject matter
- **Empowerment**: empowering the audience to question the visualization and form their own opinion in order to enable them to "shift from awareness to action" (p. 5)

When viewing visualizations, one fundamental principle is the idea of micro/macro readings. Tufte (1990) describes this as "detail [which] cumulates into larger coherent structures" (p. 37) and adds that "in all […] micro/macro designs, the same ink serves more than one informational purpose; graphical elements are multifunctioning" (p. 47).

Research Question

The data visualization found in The Economist (Figure 1) triggered interest by the project team. We observed that it does not optimally support the argument made in the article, as described above. Based on the related literature on the bill and data visualization presented in this chapter, we pursue the following research question: how can we improve the visualization found in The Economist for exploratory data analysis in the context of the article?

Methodology

In order to improve the visualization found in The Economist, we followed 3 steps: first, obtain and understand the raw data. Second, critique the existing visualization. Third, rebuild the visualization.

Obtain and understand the data

The source of the original data is given as “Chris Warshaw/The Economist” (see Figure 1). Prof. Warshaw is an associate professor at the department of political science at George Washington University, Washington, DC. Upon request, Prof. Warshaw kindly sent us the original dataset as a .csv file. It is attached as appendix 1 and shown in Figure 4.

The dataset covers major legislation polls from 1990 to 2021. It consists of 79 observations of 7 variables:

1. Identification no.
2. Name of the bill including the year
3. Year
4. Name of the polling firm
5. Percentage of respondents supporting the bill
6. Percentage of respondents opposing the bill
7. Margin of percentage of respondents supporting/opposing the bill

Each observation consists of one poll, conducted by the respective polling firm. The numbers of polls per bill vary from 2 polls (“Expand Minimum Wage (2007”) to 11 polls (“GOP Health Care Bill (2017)”).

While the purpose of the identification number is unclear and irrelevant for further analysis, all other variables are relatively transparent.

major_legislation_polls

	Bill	Year	PollingFirm	support	oppose	margin
1	Affordable Care Act (2009)	2009	ABC News/Washington Post	46	48	-2
4	Affordable Care Act (2009)	2009	CBS News/NY Times	40	45	-5
5	Affordable Care Act (2009)	2009	CNN	49	49	0
6	Affordable Care Act (2009)	2009	Fox News	38	48	-10
13	Affordable Care Act (2009)	2009	NBC News	53	43	10
15	Affordable Care Act (2009)	2009	Pew Foundation	37	46	-9
16	Affordable Care Act (2009)	2009	Quinnipiac	40	47	-7
19	American Clean Energy and Security Act (2009)	2009	ABC News/Washington Post	52	43	9
23	American Clean Energy and Security Act (2009)	2009	CNN	60	37	23

Figure 4. Head of original dataset. Source: Warshaw (2021)

We imported the .csv file into R using the read.csv function and stored it in a data frame named “polls” for further use:

```
> polls <- read.csv("major_legislation_polls.csv")
> str(polls)
'data.frame':	79 obs. of  7 variables:
 $ X          : int  1 4 5 6 13 15 16 19 23 33 ...
 $ Bill       : chr  "Affordable Care Act (2009)" "Affordable Care Act (2009)" "Affordable Care Act (2009)"
"Affordable Care Act (2009)" ...
 $ Year       : int  2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...
 $ PollingFirm: chr  "ABC News/Washington Post" "CBS News/NY Times" "CNN" "Fox News" ...
 $ support    : int  46 40 49 38 53 37 40 52 60 50 ...
 $ oppose     : int  48 45 49 48 43 46 47 43 37 39 ...
 $ margin     : int  -2 -5 0 -10 10 -9 -7 9 23 11 ...
```

Critique the existing data visualization

The data visualization in The Economist (see larger size in Figure 5) is a dotplot which prints the margin of each poll as a light blue dot, organized by the corresponding bill (depicted on the left side). Additionally, the average net support for each bill is printed as a dark blue dot. The bill discussed in the article is highlighted through bold text and a black outline on the average net support.

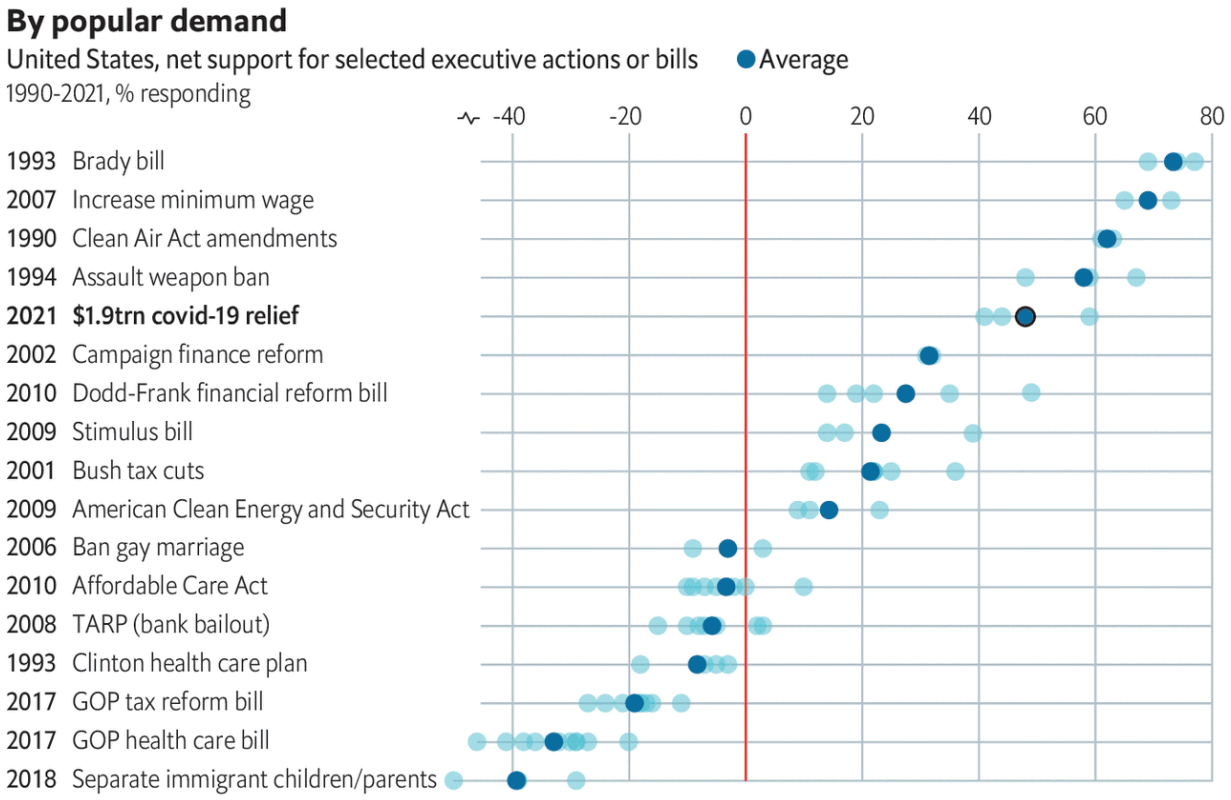


Figure 5. Original data visualization on net support for selected executive actions or bills in the U.S., larger size. Source: The Economist (2021)

The x-axis contains the net support in % of respondents of the individual polls. It ranges from -40 to +80, though some values are below -40 and displayed truncated on the left side. The value 0 is marked by a red line.

The visualization has several flaws. Most importantly, the reader is not informed about the meaning of the individual dots. In other words, there is no explanation that the individual data points are polls conducted by different firms. Rather, the reader has to extrapolate this information by considering the headline of the graph as well as the incomplete legend.

Further, the statement that 70% of the public support the bill is not actually visualized because the graph only displays net support. While the relative average popularity of the bill in comparison to other bills is still apparent, this piece of data journalism could have been done better.

Considering the fact that the article mentions the number 70 but the graph does not show it, we believe that the step *presentation transformation* (Carpendale, 2003) could be improved. The data already contain the argument, however, it is not shown in the visual presentation.

Following the principles introduced by Dörk et al. (2013), there is a lack of *disclosure* in the visualization. Since this graph has been published online, not in print, there is (in principle) unlimited space to explain the data transformation and visualization choices made when creating the graph. Additionally, a digital interactive version would allow readers to explore the data further.

Finally, a more positive remark can be made regarding the micro/macro readings of the graph (Tufte, 1990). Each bill can be read on its own, enabling the reader to explore the individual polls and their average value (micro reading). Additionally, the bills can be compared to each other in a visual way, allowing the reader to learn more about the bills' popularity in the context of other bills (macro reading).

Rebuild the visualization

Equipped with the original dataset, we embarked on recreating and improving the graph. In the following, different attempts will be presented which highlight various experiments with regards to the visualization and underlying R code.

Two final versions of the redesigned visualization will be presented in the next chapter.

$1.9trn covid-19 relief (2021)

Supporting the article (“With 70% of the public behind him, Mr Biden may not have to listen”):
Mean of support:

```
> mean(polls[34:39,"support"])
[1] 71.16667
```

The average support result **(71%)** shows that the statement in the article is true; however, only looks at % of people supporting and ignores those that oppose.

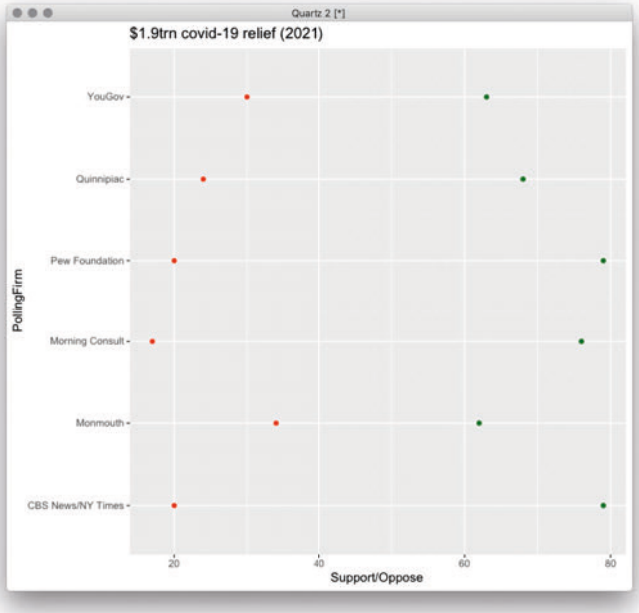
Mean of net support:

```
> mean(polls[34:39,"margin"])
[1] 47
```

The Average margin result shows that the statement in the article is still true; though the net support is only **47%** (as visualized in the graphic in The Economist)

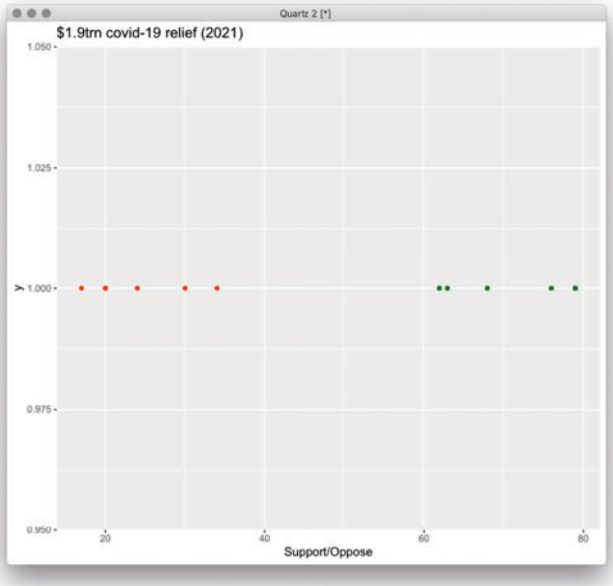
Visualize Support/Oppose by each polling firm for the Covid-19 Relief

```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("$1.9trn covid-19 relief (2021)")
```



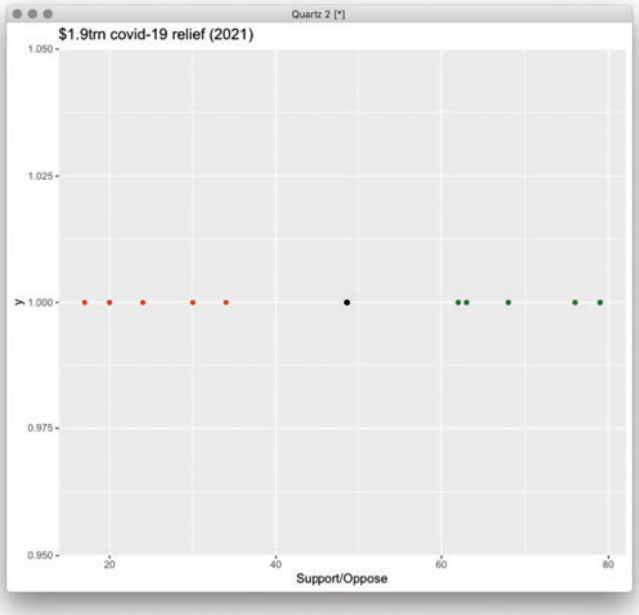
Since the data is only for Covid-19 relief, it could be reduced to one line as below. Y axis represents Covid-19 Bill (same for all values on x axis). Values on the X axis represent polling firms.

```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_point(aes(x=support,y=1),color="darkgreen") +
geom_point(aes(x=oppose,y=1),color="red") + xlab("Support/Oppose") + ggtitle("$1.9trn
covid-19 relief (2021)")
```



Covid-19 Support/Oppose graph with additional average net support value

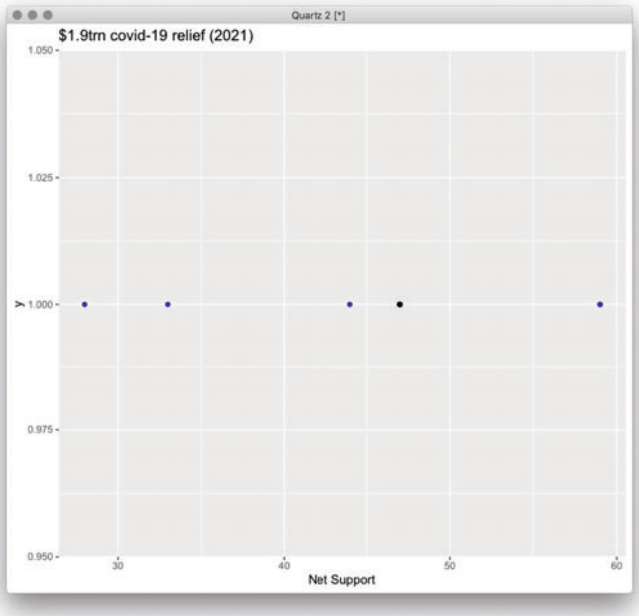
```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_point(aes(x=support,y=1),color="darkgreen") +
geom_point(aes(x=oppose,y=1),color="red") + xlab("Support/Oppose") + ggtitle("$1.9trn
covid-19 relief (2021)") + geom_point(aes(x=mean(polls[34:36,]$margin)),y=1)
```



Green = % of people who support, in different polls
Red = % of people who oppose, in different polls
Black = average (mean) of % of people who net support/oppose, across different polls

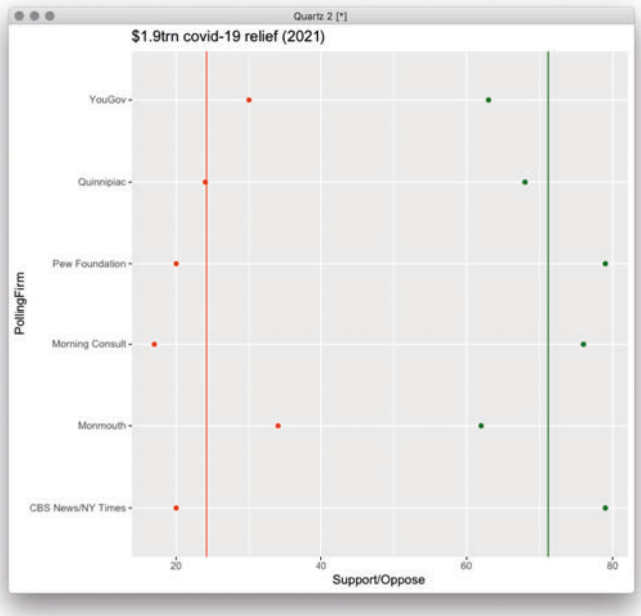
Visualize the margins and mean of margins

```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_point(aes(x=margin,y=1),color="blue") + xlab("Net Support") +
ggtitle("$1.9trn covid-19 relief (2021)") + eom_point(aes(x=mean(polls[34:39,]$margin),y=1))
```



→ this is the same as in The Economist and still makes sense; though it does not prove the point in the article (“70% of the public behind him”)
Focusing more on the individual polling firms:

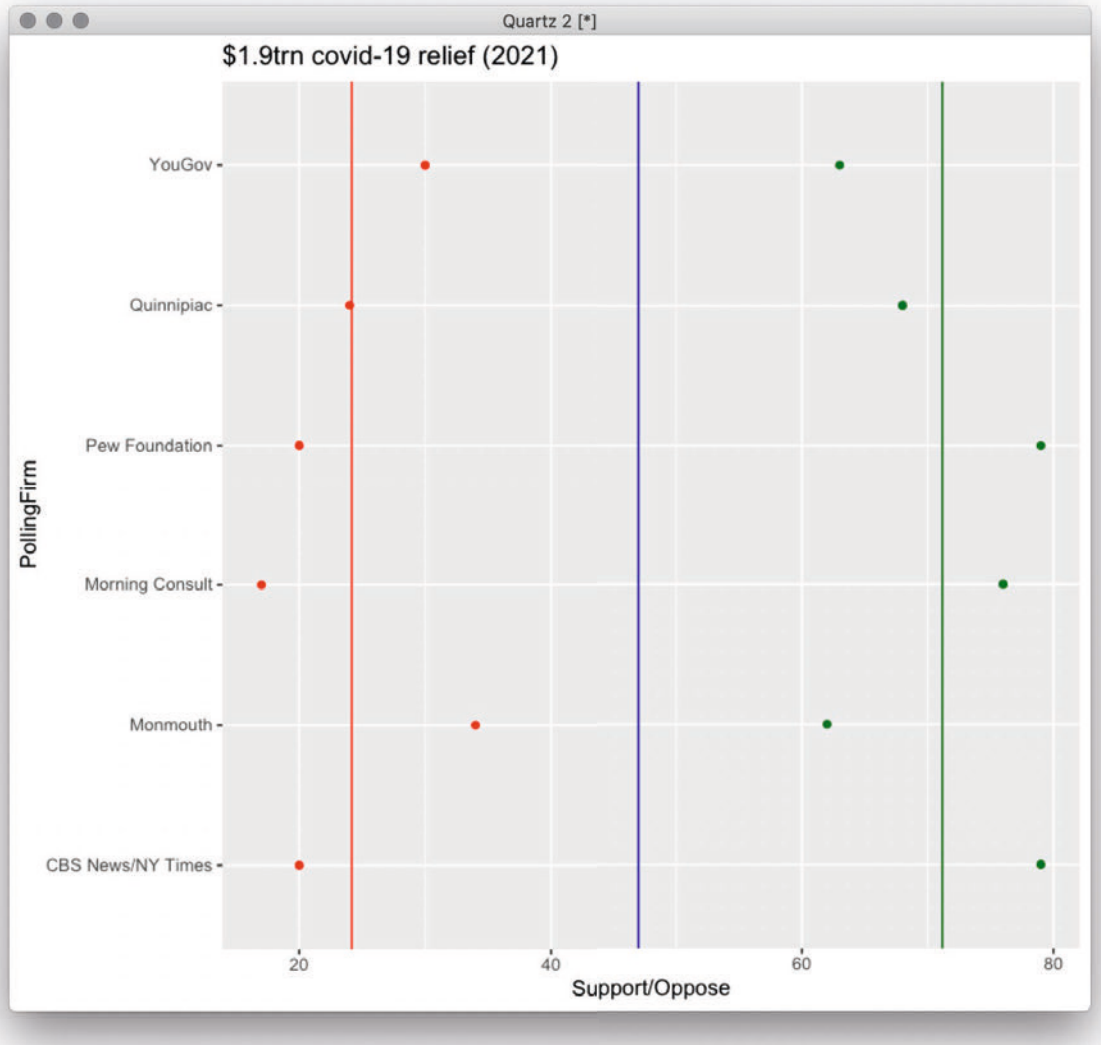
```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("$1.9trn covid-19 relief (2021)") + eom_vline(xintercept=mean(polls[34:39,"oppose"]),
color="red") + geom_vline(xintercept=mean(polls[34:39,"support"]),color="darkgreen")
```



*Lines are averages of oppose (red) and support (green)

Individual polling firms and the net support:

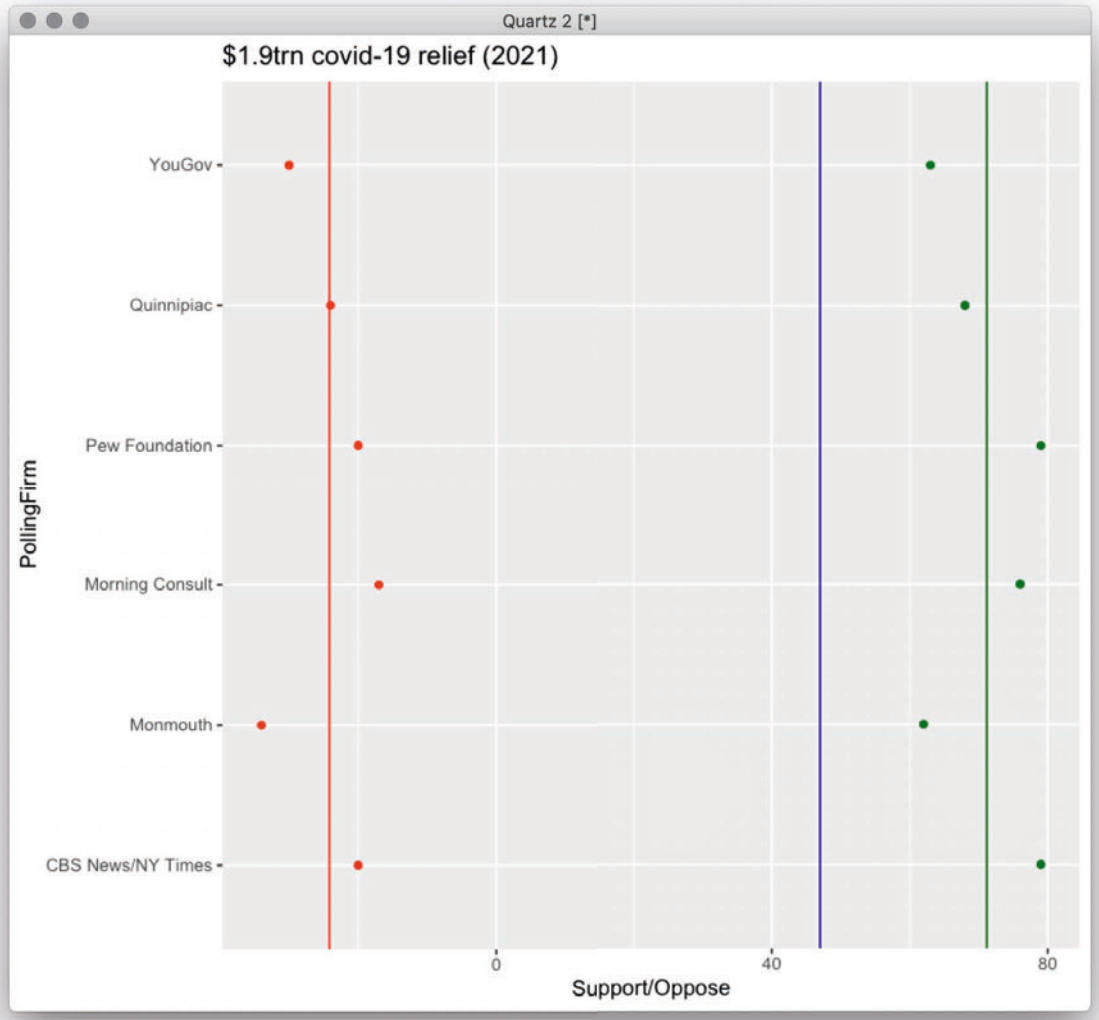
```
plot_covrelief <- ggplot(data=polls[34:39,])
meansupport_cov <- mean(polls[34:39,"support"])
meanoppose_cov <- mean(polls[34:39,"oppose"])
netsupport_cov <- meansupport_cov - meanoppose_cov
plot_covrelief + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("$1.9trn covid-19 relief (2021)") + geom_vline(xintercept=meanoppose_cov,
color="red") + geom_vline(xintercept=meansupport_cov,color="darkgreen") +
geom_vline(xintercept=netsupport_cov,color="blue")
```



*Lines are averages of oppose (red) and support (green) as well as net support (blue)

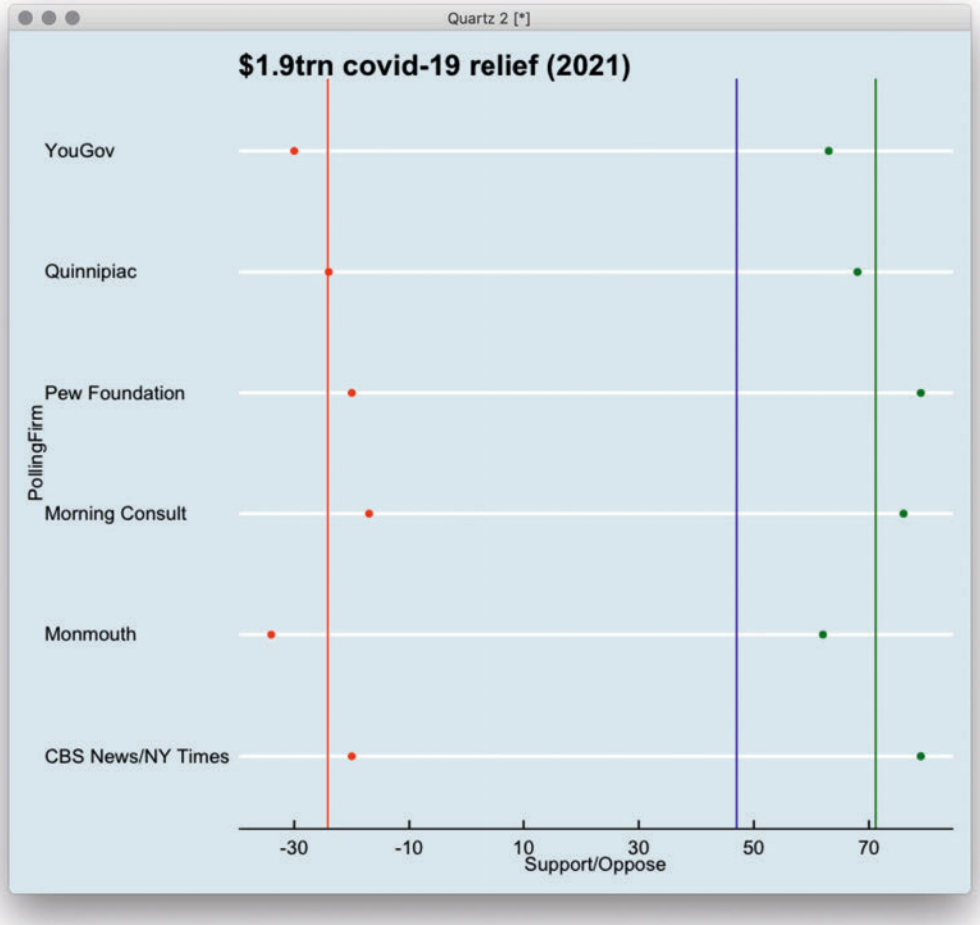
The “oppose” should be taken as negative, then the graph makes more sense:
→ this graph includes the average net support (47%) and show the average support (71%)

```
plot_covrelief <- ggplot(data=polls[34:39,])
meansupport_cov <- mean(polls[34:39,"support"])
meanoppose_cov <- -mean(polls[34:39,"oppose"])
netsupport_cov <- meansupport_cov + meanoppose_cov
plot_covrelief + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=-oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("$1.9trn covid-19 relief (2021)") + geom_vline(xintercept=meanoppose_cov,
color="red") + geom_vline(xintercept=meansupport_cov,color="darkgreen") +
geom_vline(xintercept=netsupport_cov,color="blue")
```



Adding more frequent values on x axis and changing the color to blue make the graph easier to read

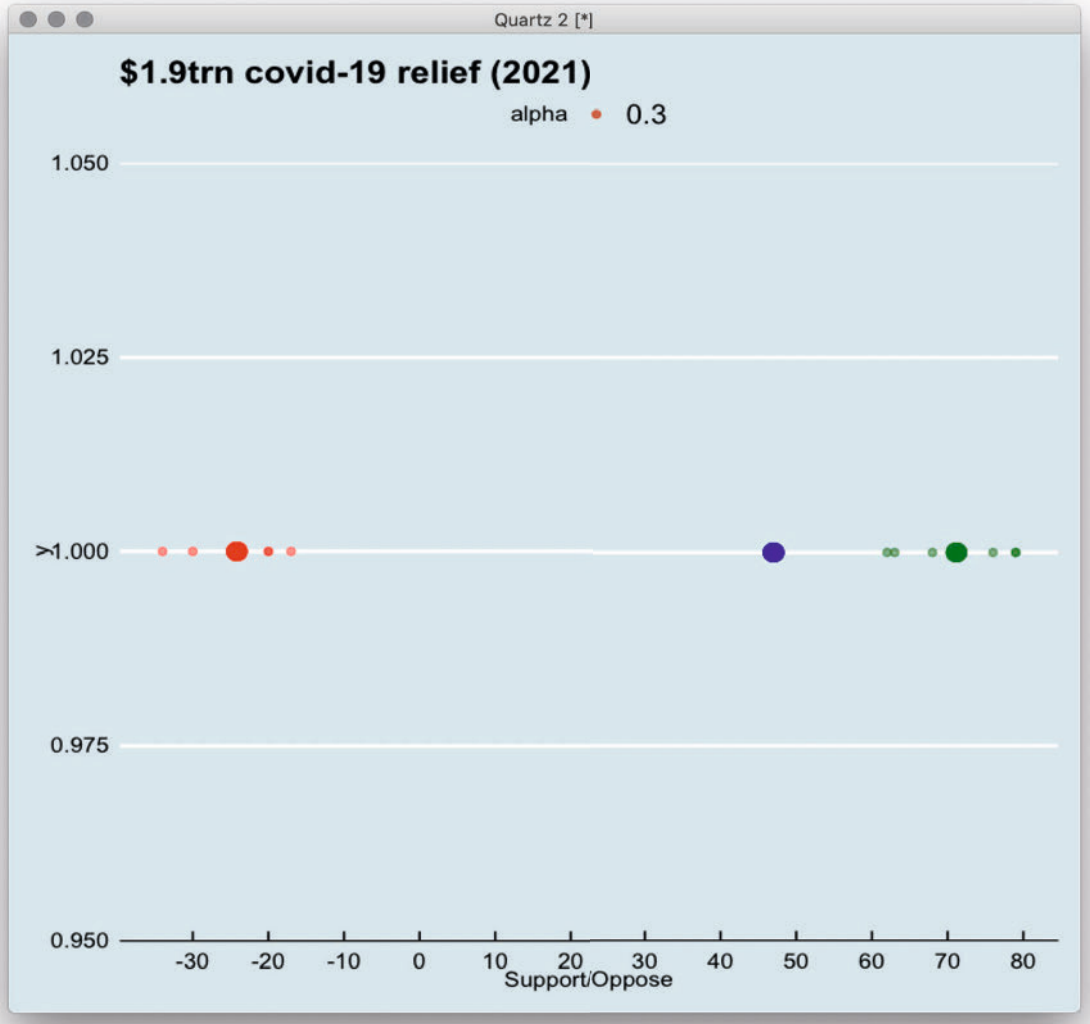
```
plot_covrelief <- ggplot(data=polls[34:39,])
meansupport_cov <- mean(polls[34:39,"support"])
meanoppose_cov <- -mean(polls[34:39,"oppose"])
netsupport_cov <- meansupport_cov + meanoppose_cov
plot_covrelief + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=-oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("$1.9trn covid-19 relief (2021)") + geom_vline(xintercept=meanoppose_cov,
color="red") + geom_vline(xintercept=meansupport_cov,color="darkgreen") +
geom_vline(xintercept=netsupport_cov,color="blue") + theme_economist() +
scale_x_continuous(breaks=seq(-50,100,by=20))
```



*Lines are averages of oppose (red) and average of support (green) as well as net support (blue)

Again, reduced to one line and with larger averages (mean) of support, oppose, and net support, as well as transparency (alpha) of dots:

```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_point(aes(x=support,y=1, alpha=0.3),color="darkgreen") +
geom_point(aes(x=-oppose,y=1, alpha=0.3),color="red") + xlab("Support/Oppose") +
ggtitle("$1.9trn covid-19 relief (2021)") + geom_point(aes(x=meanoppose_cov,y=1), size=4,
color="red") + geom_point(aes(x=meansupport_cov, y=1), size=4, color="darkgreen") +
geom_point(aes(x=netsupport_cov,y=1),size=4,color="blue") + theme_economist() +
scale_x_continuous(breaks=seq(-50,100,by=10))
```

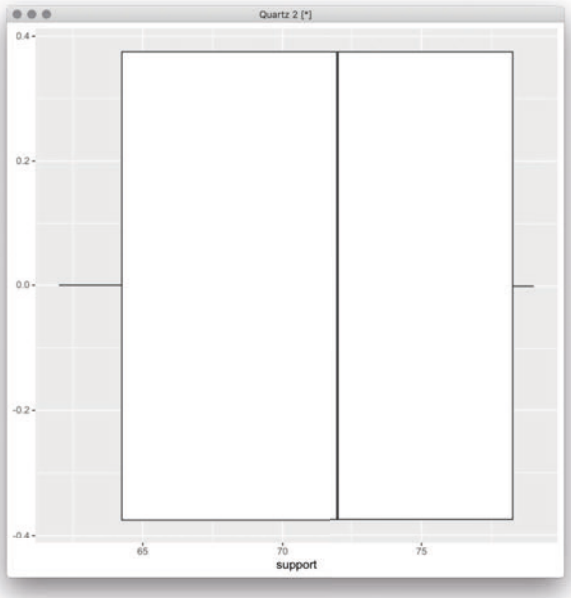


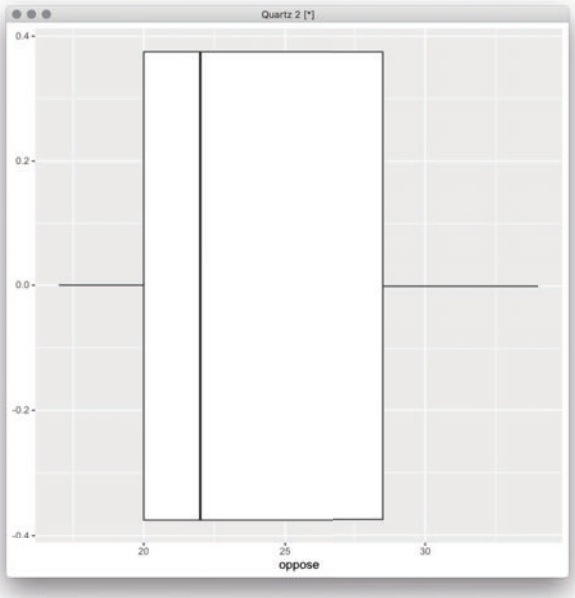
Boxplot

(note that the boxplot uses median instead of mean)

Boxplot of support (left) and oppose (right)

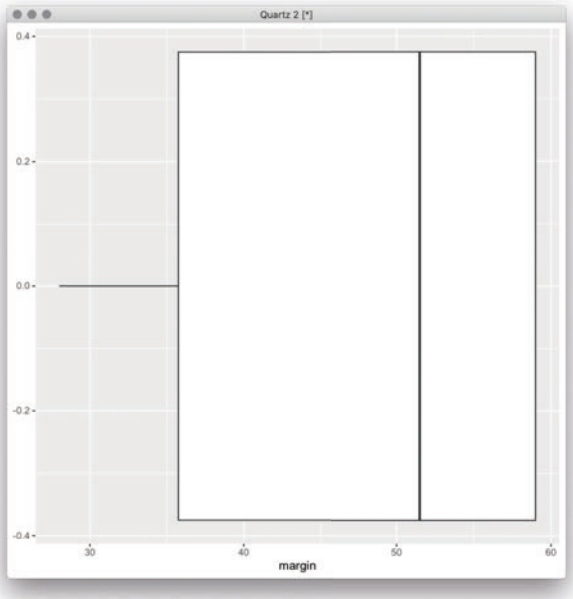
```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_boxplot(aes(x=support))
plot_covrelief + geom_boxplot(aes(x=oppose))
```





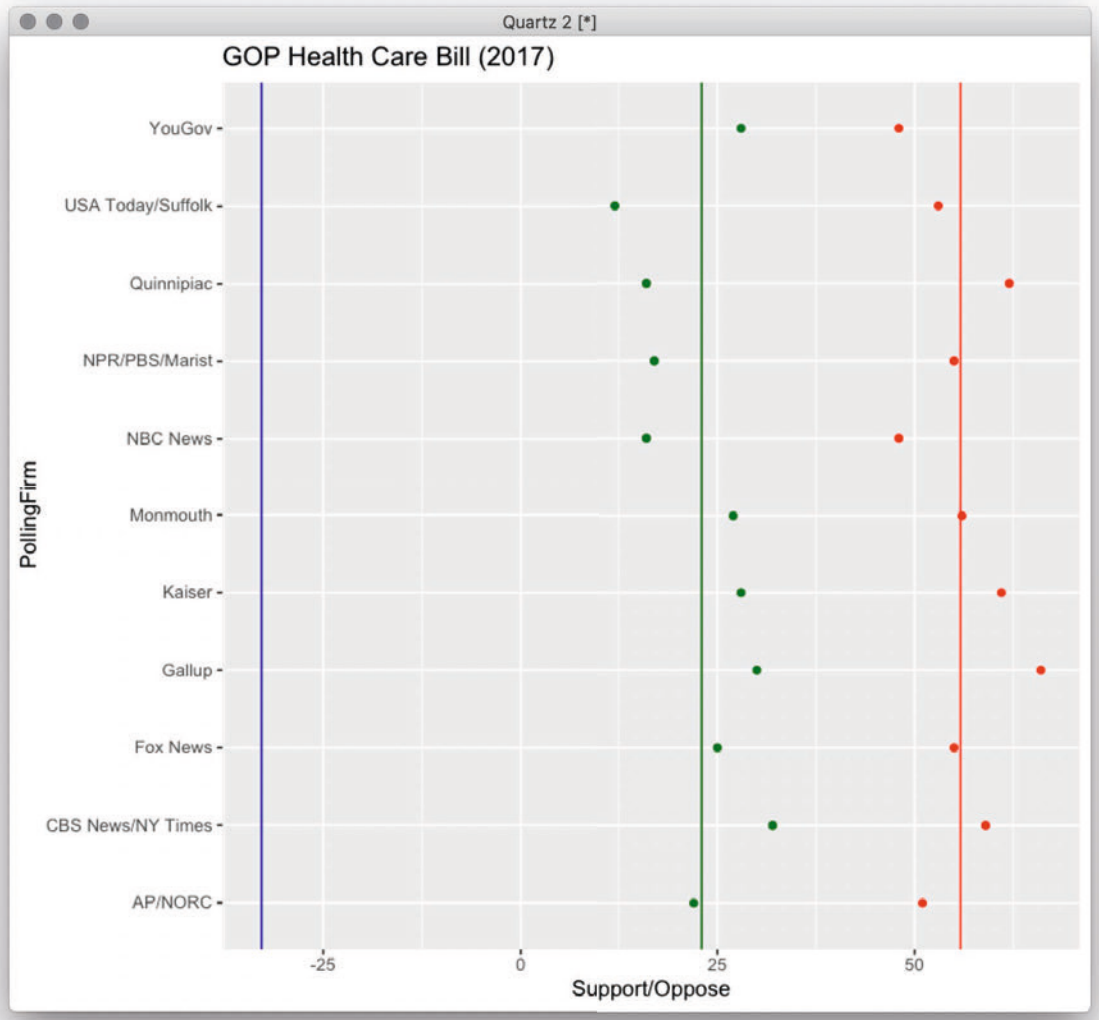
Boxplot of margin

```
plot_covrelief <- ggplot(data=polls[34:39,])
plot_covrelief + geom_boxplot(aes(x=margin))
```



GOP Health Care Bill (2017)

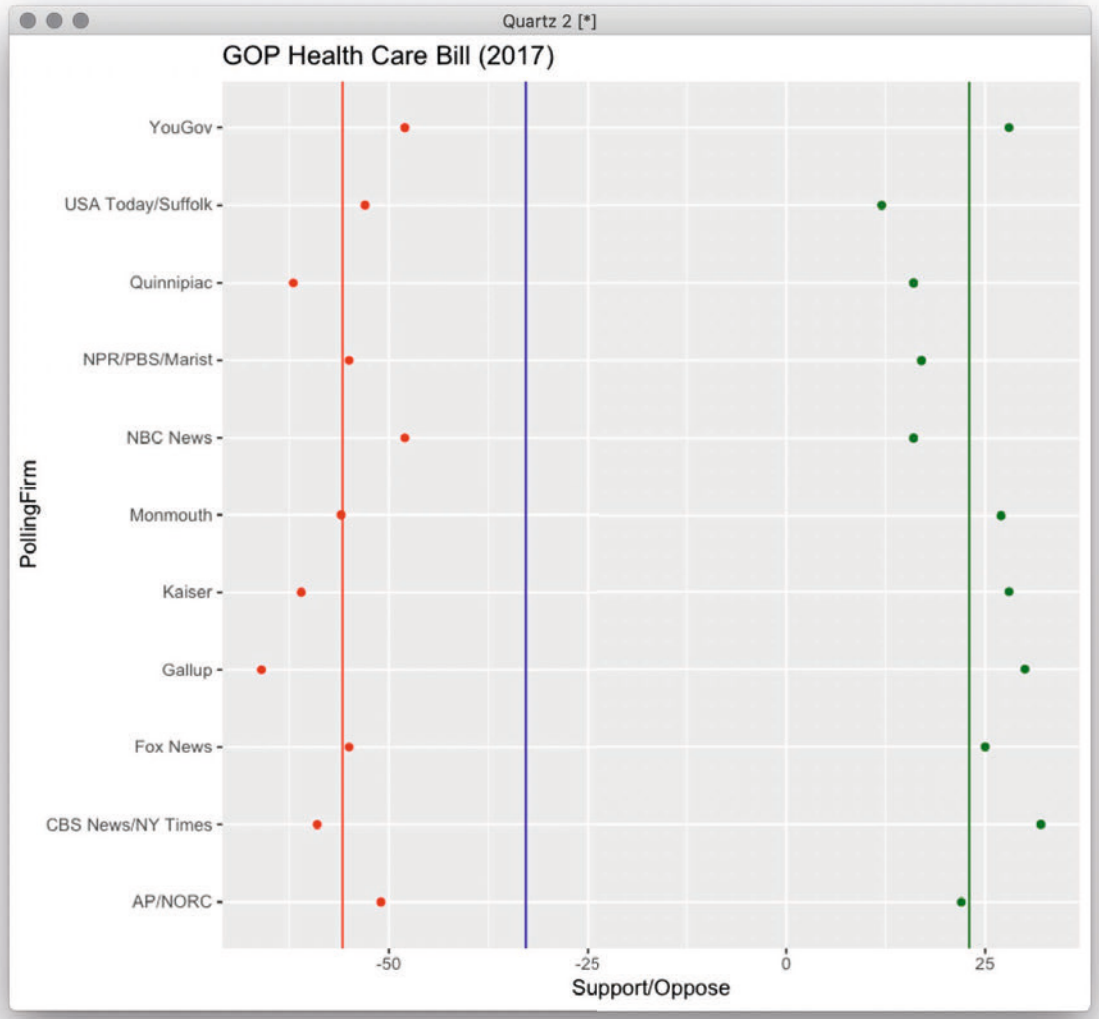
```
plot_GOPhealthcare <- ggplot(data=polls[47:57,])
meansupport_GOPhealthcare <- mean(polls[47:57,"support"])
meanoppose_GOPhealthcare <- mean(polls[47:57,"oppose"])
netsupport_GOPhealthcare <- meansupport_GOPhealthcare - meanoppose_GOPhealthcare
plot_GOPhealthcare + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("GOP Health Care Bill (2017)") +
geom_vline(xintercept=meanoppose_GOPhealthcare, color="red") +
geom_vline(xintercept=meansupport_GOPhealthcare,color="darkgreen") +
geom_vline(xintercept=netsupport_GOPhealthcare,color="blue")
```



*Lines are averages of oppose (red) and support (green) as well as net support (blue)
Taking oppose as negative values results in a better graph:

```
plot_GOPhealthcare <- ggplot(data=polls[47:57,])
```

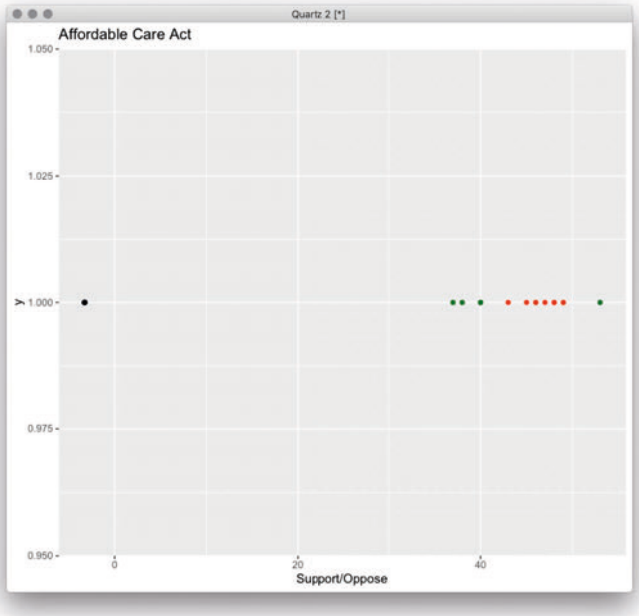
```
meansupport_GOPhealthcare <- mean(polls[47:57,"support"])
meanoppose_GOPhealthcare <- -mean(polls[47:57,"oppose"])
netsupport_GOPhealthcare <- meansupport_GOPhealthcare +
meanoppose_GOPhealthcare
plot_GOPhealthcare + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=-oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("GOP Health Care Bill (2017)") +
geom_vline(xintercept=meanoppose_GOPhealthcare, color="red") +
geom_vline(xintercept=meansupport_GOPhealthcare,color="darkgreen") +
geom_vline(xintercept=netsupport_GOPhealthcare,color="blue")
```



Affordable Care Act (2009)

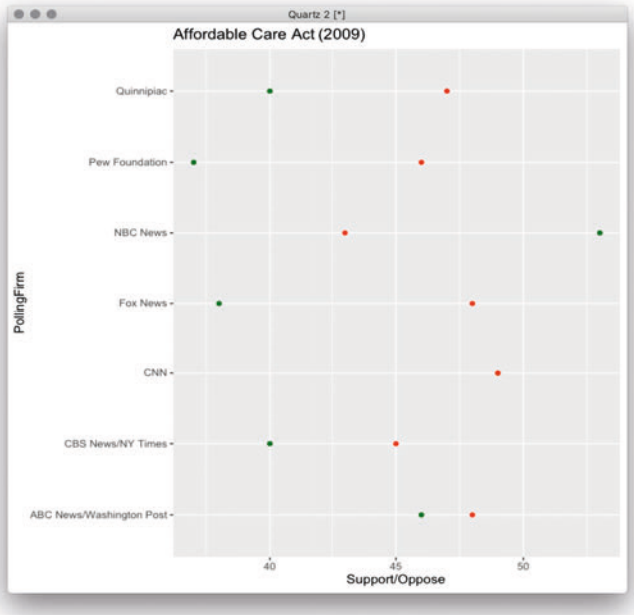
```
plot_afcaact <- ggplot(data=polls[1:7,])
plot_afcaact + geom_point(aes(x=support,y=1),color="darkgreen") +
```

```
geom_point(aes(x=oppose,y=1),color="red") + xlab("Support/Oppose") + ggtitle("Affordable Care Act") + geom_point(aes(x=mean(polls[1:7,]$margin)),y=1)
```



Visualize Support/Oppose for one Bill only (for Affordable Care Act (2009))

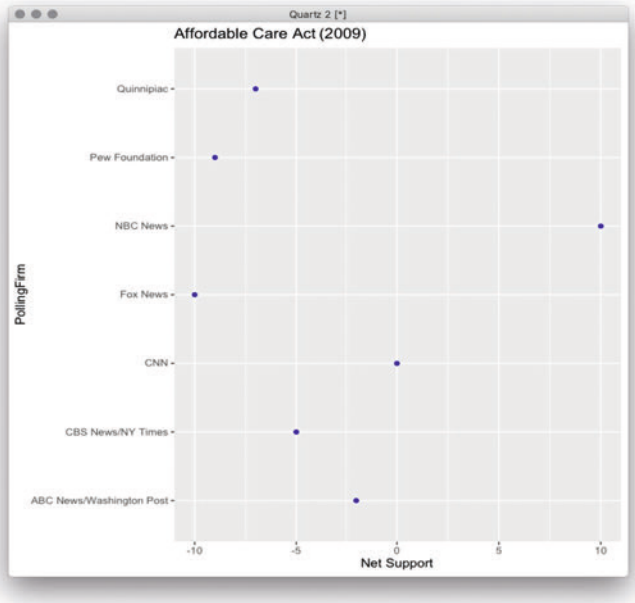
```
plot_afcaact <- ggplot(data=polls[1:7,])
plot_afcaact + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Affordable Care Act (2009)")
```



Green = support, Red = oppose

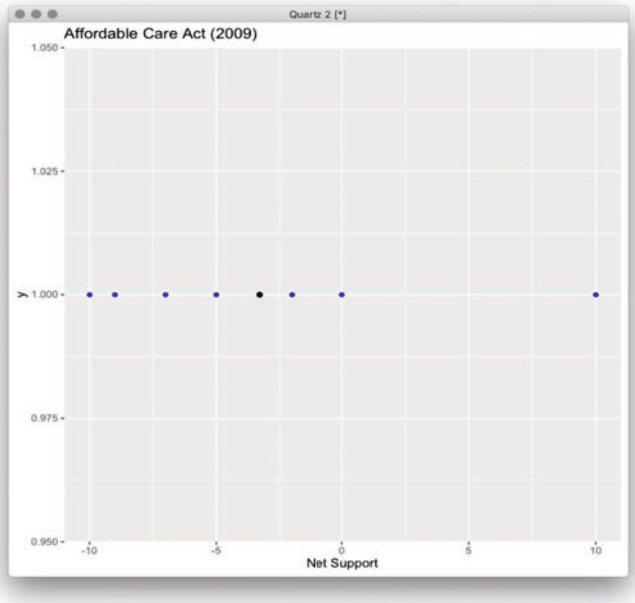
Visualizing net support (margin)

```
plot_afcaact + geom_point(aes(x=margin,y=PollingFirm),color="blue") + xlab("Net Support")
+ ggtitle("Affordable Care Act (2009)")
```



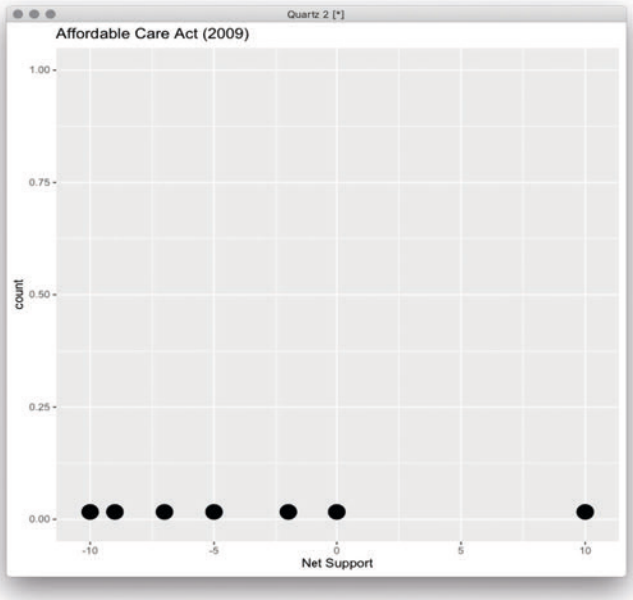
Margin of all polling firms (blue) and mean (black) for Affordable Care Act in one line

```
plot_afcaact + geom_point(aes(x=margin,y=1),color="blue") + xlab("Net Support") +
ggtitle("Affordable Care Act (2009)") + geom_point(aes(x=mean(polls[1:7,]$margin),y=1))
```



Create a dotplot of the margin of all polling firms

```
plot_afcaact + geom_dotplot(aes(x=margin)) + xlab("Net Support") + ggtitle("Affordable Care Act (2009)")
```



Bush Tax Cuts (2001)

How popular were Bush Tax Cuts 2001?

Mean of support:

```
> mean(polls[19:25,"support"])
[1] 56.57143
```

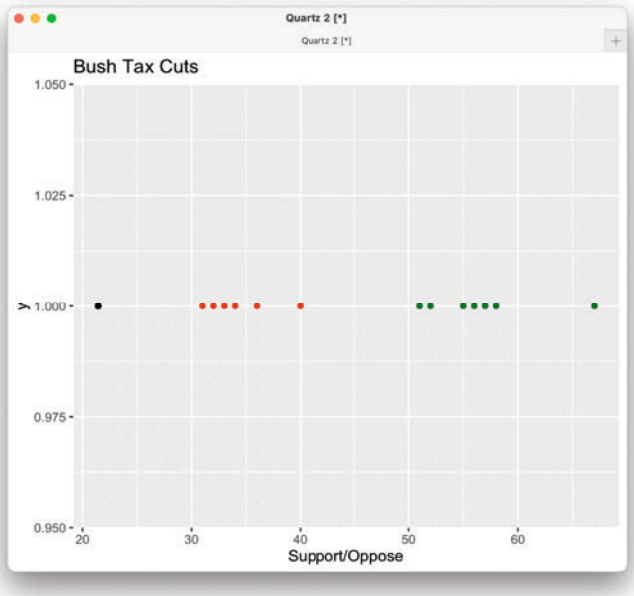
-> Bush’s tax cut wasn’t really supported by the public according to the data. Only 56,6% people supported this Tax Cut. In comparison, Biden's Covid-19 relief is far more popular.

Mean of net support:

```
> mean(polls[19:25,"margin"])
[1] 21.42857
```

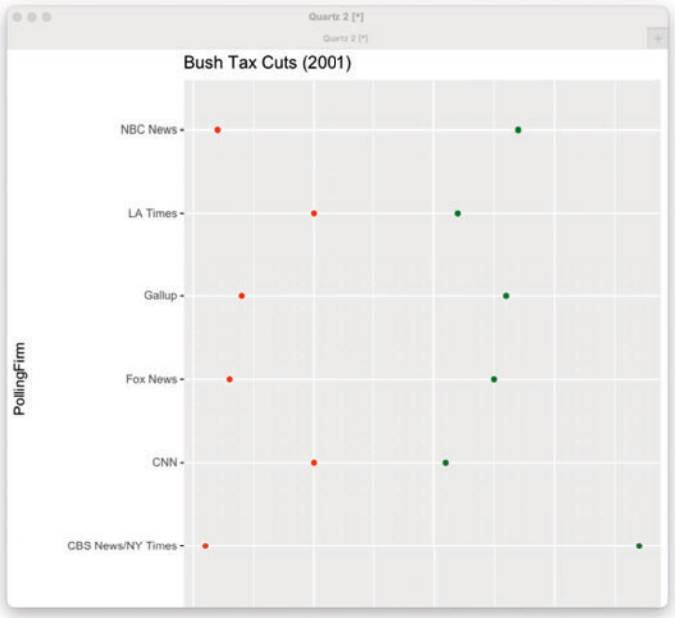
-> net support is really low..

```
plot_BushTC + geom_point(aes(x=support,y=1),color="darkgreen") +
geom_point(aes(x=oppose,y=1),color="red") + xlab("Support/Oppose") + ggtitle("Bush Tax
Cuts") + geom_point(aes(x=mean(polls[19:25,]$margin)),y=1)
```



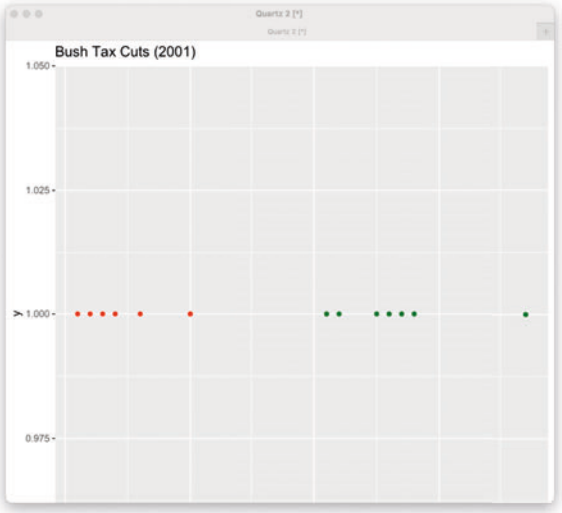
Visualize Support/Oppose by each polling firm

```
plot_BushTC<- ggplot(data=polls[19:25,])
plot_BushTC+ geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Bush Tax Cuts (2001)")
```



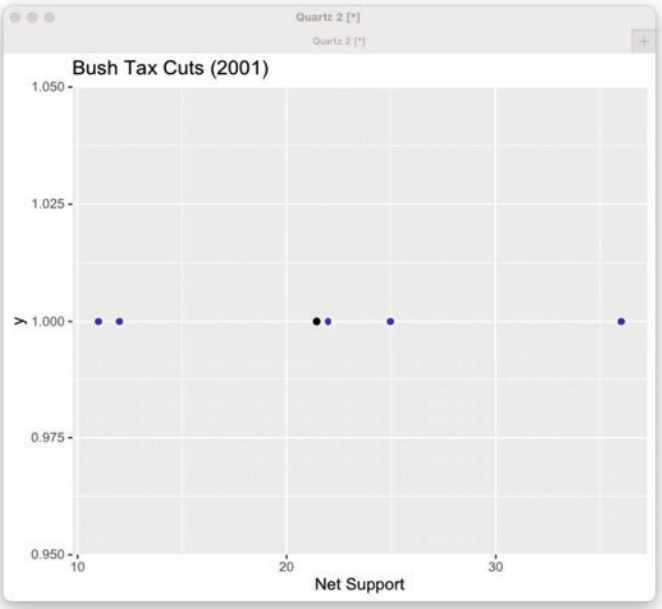
Reduced to one line:

```
plot_BushTC + geom_point(aes(x=support,y=1),color="darkgreen") +
geom_point(aes(x=oppose,y=1),color="red") + xlab("Support/Oppose") + ggtitle("Bush Tax
Cuts (2001)")
```



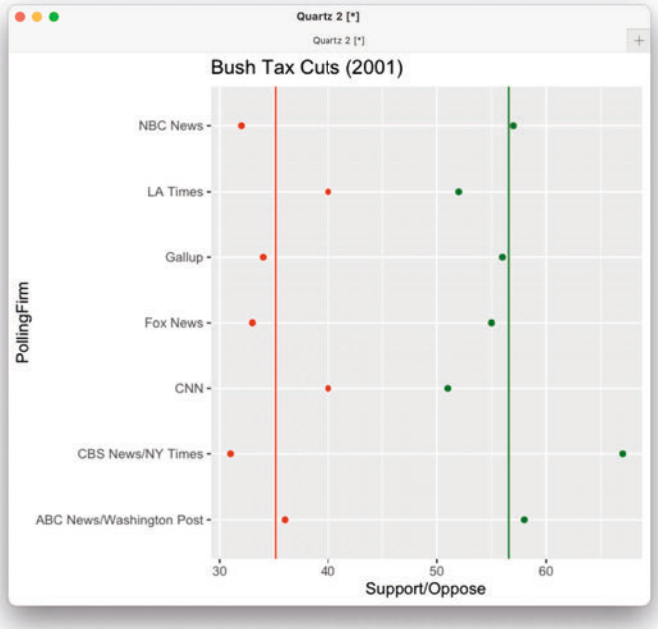
Visualize the margins and mean of margins

```
plot_BushTC + geom_point(aes(x=margin,y=1),color="blue") + xlab("Net Support") +
ggtitle("Bush Tax Cuts (2001)") + geom_point(aes(x=mean(polls[19:25,]$margin),y=1))
```



Focusing more on the individual polling firms:

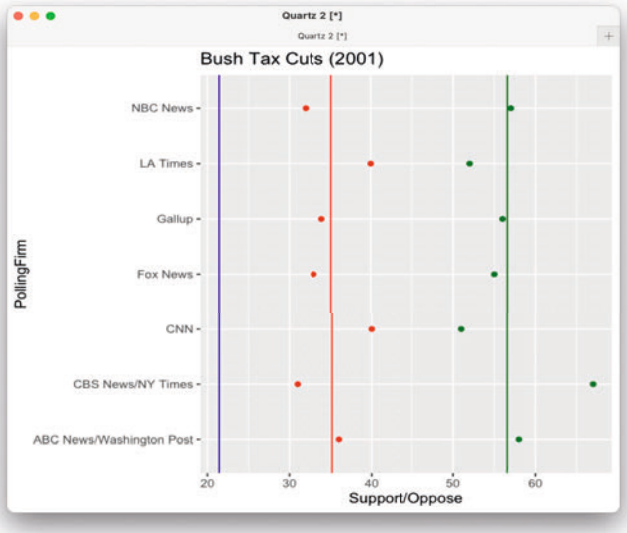
```
plot_BushTC + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Bush Tax Cuts (2001)") + geom_vline(xintercept=mean(polls[19:25,"oppose"]),
color="red") + geom_vline(xintercept=mean(polls[19:25,"support"]),color="darkgreen")
```



*Lines are averages of oppose (red) and support (green)

And with the net support:

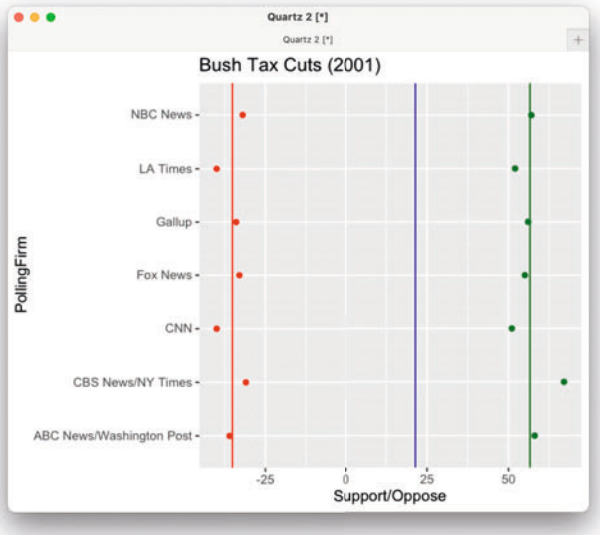
```
plot_BushTC<-ggplot(data=polls[19:25,])
meansupport_btc <- mean(polls[19:25,"support"])
meanoppose_btc <- mean(polls[19:25,"oppose"])
netsupport_btc <- meansupport_btc - meanoppose_btc
plot_BushTC + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Bush Tax Cuts (2001)") + geom_vline(xintercept=meanoppose_btc, color="red") +
geom_vline(xintercept=meansupport_btc,color="darkgreen") +
geom_vline(xintercept=netsupport_btc,color="blue")
```



*Lines are averages of oppose (red) and support (green) as well as net support (blue)

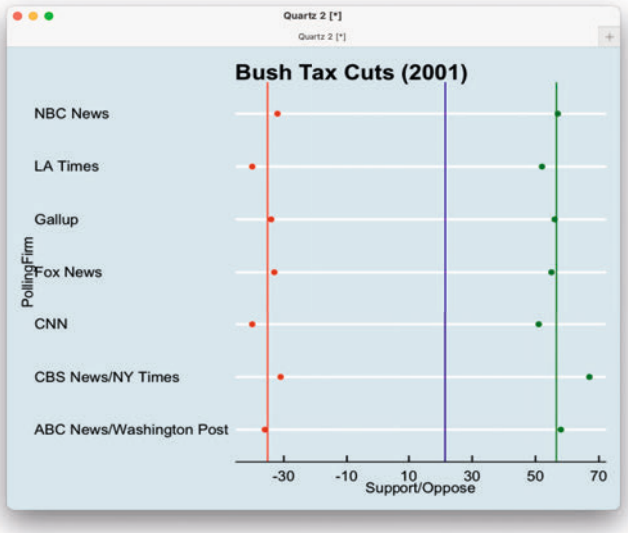
The “oppose” should be taken as negative
Bush’s Tax Cuts: net support (21%), average support (57%)

```
plot_BushTC<-ggplot(data=polls[19:25,])
meansupport_btc <- mean(polls[19:25,"support"])
meanoppose_btc <- -mean(polls[19:25,"oppose"])
netsupport_btc <- meansupport_btc + meanoppose_btc
plot_BushTC + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=-oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Bush Tax Cuts (2001)") + geom_vline(xintercept=meanoppose_btc, color="red") +
geom_vline(xintercept=meansupport_btc,color="darkgreen") +
geom_vline(xintercept=netsupport_btc,color="blue")
```



A bit “prettier” (using the theme) in the following:

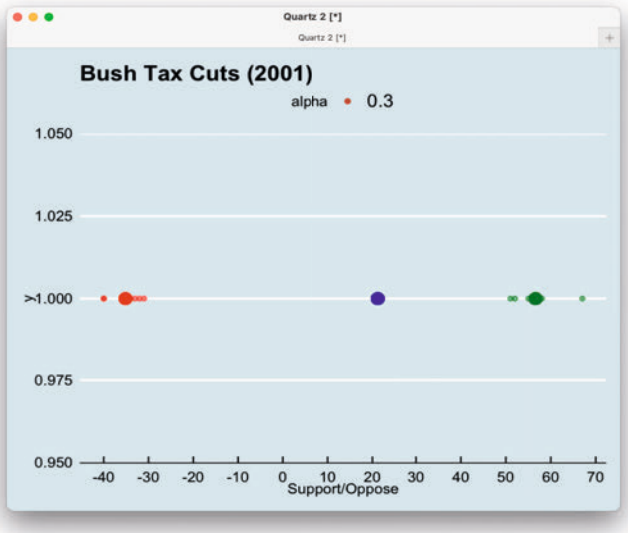
```
plot_BushTC<-ggplot(data=polls[19:25,])
meansupport_btc <- mean(polls[19:25,"support"])
meanoppose_btc <- -mean(polls[19:25,"oppose"])
netsupport_btc <- meansupport_btc + meanoppose_btc
plot_BushTC + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=-oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Bush Tax Cuts (2001)") + geom_vline(xintercept=meanoppose_btc, color="red") +
geom_vline(xintercept=meansupport_btc,color="darkgreen") +
geom_vline(xintercept=netsupport_btc,color="blue") + theme_economist() +
scale_x_continuous(breaks=seq(-50,100,by=20))
```



*Lines are averages of oppose (red) and average of support (green) as well as net support (blue)

Again, reduced to one line and with larger averages (mean) of support, oppose, and net support, as well as transparency (alpha) of dots:

```
plot_BushTC <- ggplot(data=polls[19:25,])
plot_BushTC + geom_point(aes(x=support,y=1, alpha=0.3),color="darkgreen") +
geom_point(aes(x=-oppose,y=1, alpha=0.3),color="red") + xlab("Support/Oppose") +
ggtitle("Bush Tax Cuts (2001)") + geom_point(aes(x=meanoppose_btc,y=1), size=4,
color="red") + geom_point(aes(x=meansupport_btc, y=1), size=4, color="darkgreen") +
geom_point(aes(x=netsupport_btc,y=1),size=4,color="blue") + theme_economist() +
scale_x_continuous(breaks=seq(-50,100,by=10))
```



Clinton Health Care Plan (1993)

How popular was the Clinton Health Care Plan 1993? → Very unpopular

Mean of support:

```
> mean(polls[30:33,"support"])
[1] 40
```

-> Clinton Health Care Plan 1993 wasn't really supported by the public according to the data. Only 40% of people supported Clinton's Health Care Plan. In comparison, Biden's Covid-19 relief is far more popular.

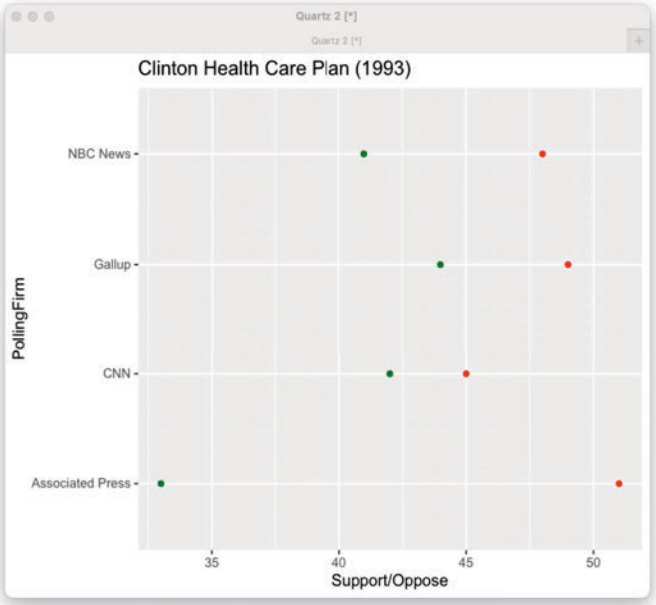
Mean of net support:

```
> mean(polls[30:33,"margin"])
[1] -8.25
```

-> More people are opposing the Plan than supporting it

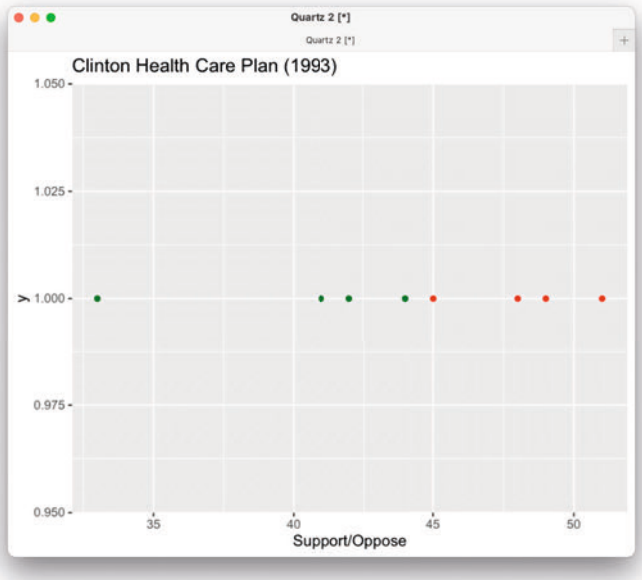
Visualize Support/Oppose by each polling firm

```
plot_clintonhc<- ggplot(data=polls[30:33,])
plot_clintonhc+ geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Clinton Health Care Plan (1993)")
```



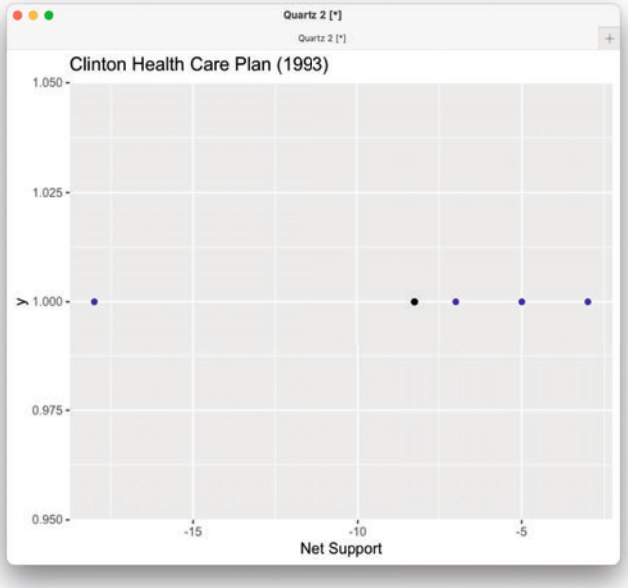
Reduced to one line:

```
plot_BushTC + geom_point(aes(x=support,y=1),color="darkgreen") +
geom_point(aes(x=oppose,y=1),color="red") + xlab("Support/Oppose") + ggtitle("$1.9trn
covid-19 relief (2021)")
```



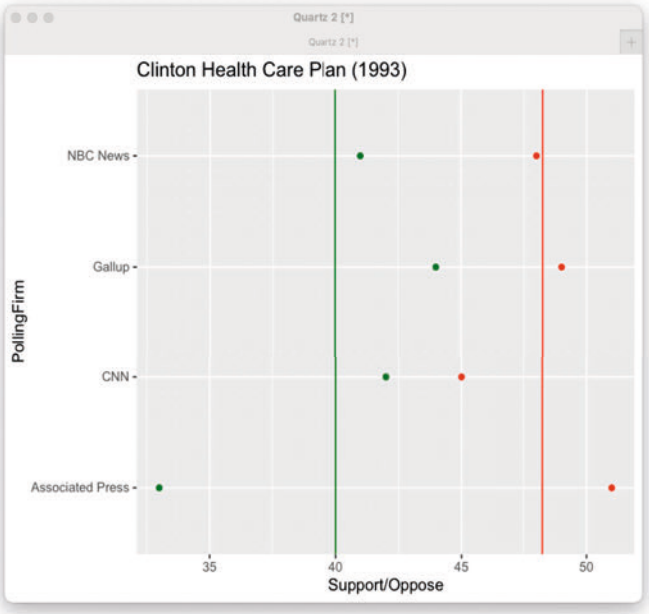
Visualize the margins and mean of margins

```
plot_clintonhc + geom_point(aes(x=margin,y=1),color="blue") + xlab("Net Support") +
ggtitle("Clinton Health Care Plan (1993)") +
geom_point(aes(x=mean(polls[30:33,]$margin),y=1))
```



Focusing more on the individual polling firms:

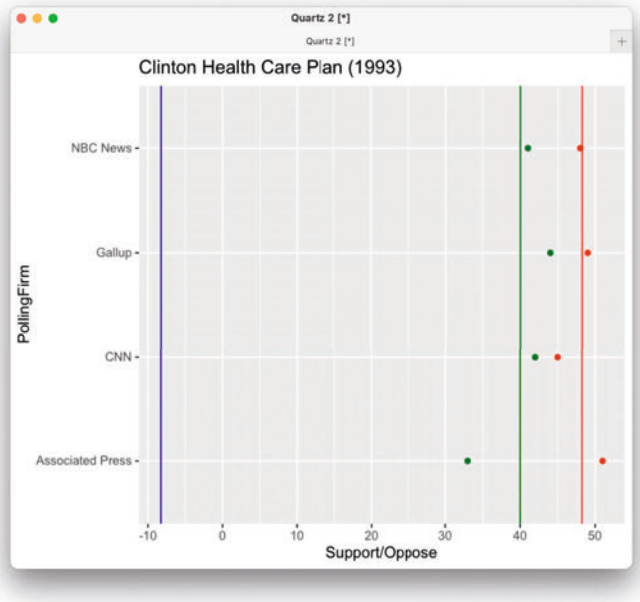
```
plot_clintonhc + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Clinton Health Care Plan (1993)") +
geom_vline(xintercept=mean(polls[30:33,"oppose"]), color="red") +
geom_vline(xintercept=mean(polls[30:33,"support"]),color="darkgreen")
```



Lines are averages of oppose (red) and support (green)

And with the net support:

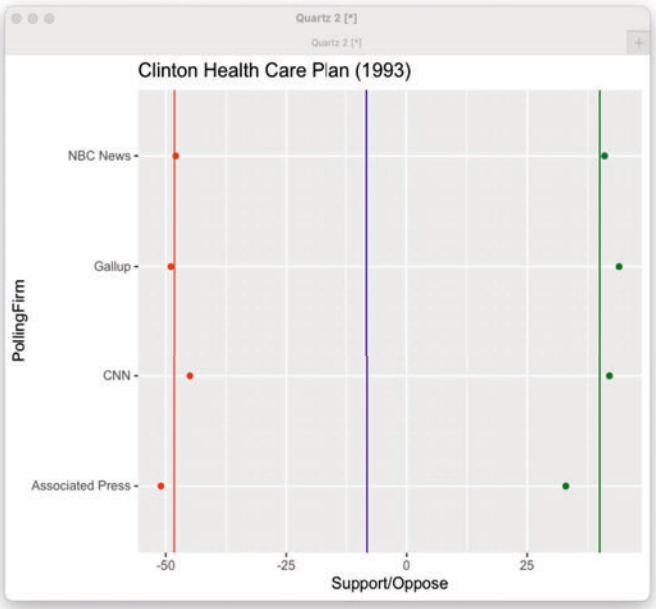
```
plot_clintonhc<-ggplot(data=polls[30:33,])
meansupport_chc <- mean(polls[30:33,"support"])
meanoppose_chc <- mean(polls[30:33,"oppose"])
netsupport_chc <- meansupport_chc - meanoppose_chc
plot_clintonhc + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Clinton Health Care Plan (1993)") + geom_vline(xintercept=meanoppose_chc,
color="red") + geom_vline(xintercept=meansupport_chc,color="darkgreen") +
geom_vline(xintercept=netsupport_chc,color="blue")
```



Lines are averages of oppose (red) and support (green) as well as net support (blue) (Negative)

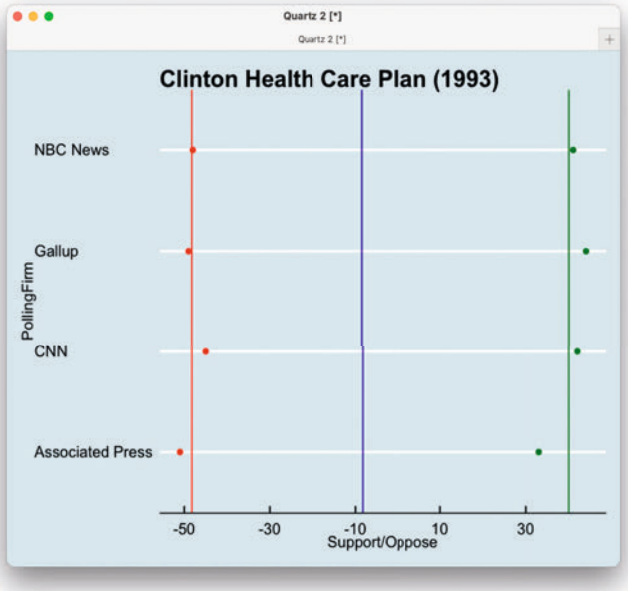
The “oppose” should be taken as negative
Biden’s graph: net support (47%) and show the average support (71%)

```
plot_clintonhc<- ggplot(data=polls[30:33,])
meansupport_chc <- mean(polls[30:33,"support"])
meanoppose_chc <- -mean(polls[30:33,"oppose"])
netsupport_chc <- meansupport_chc + meanoppose_chc
plot_clintonhc + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=-oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Clinton Health Care Plan (1993)") + geom_vline(xintercept=meanoppose_chc,
color="red") + geom_vline(xintercept=meansupport_chc,color="darkgreen") +
geom_vline(xintercept=netsupport_chc,color="blue")
```



A bit “prettier” (using the theme) in the following:

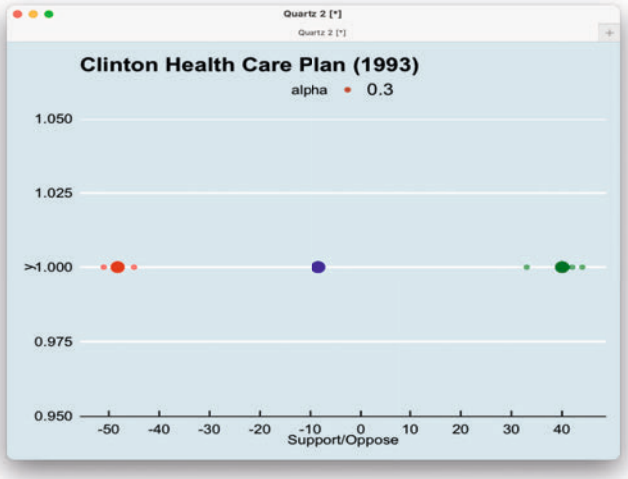
```
plot_clintonhc<- ggplot(data=polls[30:33,])
meansupport_chc <- mean(polls[30:33,"support"])
meanoppose_chc <- -mean(polls[30:33,"oppose"])
netsupport_chc <- meansupport_chc + meanoppose_chc
plot_clintonhc + geom_point(aes(x=support,y=PollingFirm),color="darkgreen") +
geom_point(aes(x=-oppose,y=PollingFirm),color="red") + xlab("Support/Oppose") +
ggtitle("Clinton Health Care Plan (1993)") + geom_vline(xintercept=meanoppose_chc,
color="red") + geom_vline(xintercept=meansupport_chc,color="darkgreen") +
geom_vline(xintercept=netsupport_chc,color="blue") +  theme_economist() +
scale_x_continuous(breaks=seq(-50,100,by=20))
```



Lines are averages of oppose (red) and average of support (green) as well as net support (blue)

Again, reduced to one line and with larger averages (mean) of support, oppose, and net support, as well as transparency (alpha) of dots:

```
plot_clintonhc<- ggplot(data=polls[30:33,])
 plot_clintonhc + geom_point(aes(x=support,y=1, alpha=0.3),color="darkgreen") +
geom_point(aes(x=-oppose,y=1, alpha=0.3),color="red") + xlab("Support/Oppose") +
ggtitle("Clinton Health Care Plan (1993)") + geom_point(aes(x=meanoppose_chc,y=1),
size=4, color="red") + geom_point(aes(x=meansupport_chc, y=1), size=4,
color="darkgreen") + geom_point(aes(x=netsupport_chc,y=1),size=4,color="blue") +
theme_economist() + scale_x_continuous(breaks=seq(-50,100,by=10))
```



Generic Function to display dotplot based on row numbers

Function generates a dotplot for any given bill. It requires a start and end row from the polls data set, as well as a title for the chart (see examples below).

Requires:

- data frame named “polls” containing data from the CSV file from C. Warshaw
- ggplot2
- ggthemes

```
display_dots <-function(temp_startrow,temp_endrow,temp_billtitle) {
        plot_genericbill <- ggplot(data=polls[temp_startrow:temp_endrow,])
        meansupport_genericbill <- mean(polls[temp_startrow:temp_endrow,"support"])
        meanoppose_genericbill <- -mean(polls[temp_startrow:temp_endrow,"oppose"])
        netsupport_genericbill <- meansupport_genericbill + meanoppose_genericbill

        plot_genericbill + geom_point(aes(x=support,y=1, alpha=0.3),color="darkgreen") +
geom_point(aes(x=-oppose,y=1, alpha=0.3),color="red") + xlab("Support/Oppose") +
ggtitle(temp_billtitle) + geom_point(aes(x=meanoppose_genericbill,y=1, alpha=0.3), size=4,
color="red") + geom_point(aes(x=meansupport_genericbill, y=1, alpha=0.3), size=4,
color="darkgreen") + geom_point(aes(x=netsupport_genericbill,y=1),size=4,color="blue") +
theme_economist() + xlim(-100,100)
}
```

Examples:

```
> display_dots(34,39,"Covid")
> display_dots(47,57,"GOP Healthcare Bill")
```

Generic Function to display dotplot based on row numbers, choose y-value and add title

Function generates a dotplot for any given bill. It requires a start and end row from the polls data set, as well as a title for the chart (see examples below) and a y-row.

Requires:

- data frame named “polls” containing data from the CSV file from C. Warshaw
- ggplot2
- ggthemes

```
display_dots <-function(temp_startrow,temp_endrow,temp_billtitle,y_row) {
        plot_genericbill <- ggplot(data=polls[temp_startrow:temp_endrow,])
        meansupport_genericbill <- mean(polls[temp_startrow:temp_endrow,"support"])
        meanoppose_genericbill <- -mean(polls[temp_startrow:temp_endrow,"oppose"])
        netsupport_genericbill <- meansupport_genericbill + meanoppose_genericbill

        plot_genericbill + geom_point(aes(x=support,y=y_row, alpha=0.3),color="darkgreen")
+ geom_point(aes(x=-oppose,y=y_row, alpha=0.3),color="red") + xlab("Support/Oppose") +
```

```
ggtitle(temp_billtitle) + geom_point(aes(x=meanoppose_genericbill,y=y_row, alpha=0.3),
size=4, color="red") + geom_point(aes(x=meansupport_genericbill, y=y_row, alpha=0.3),
size=4, color="darkgreen") +
geom_point(aes(x=netsupport_genericbill,y=y_row),size=4,color="blue") +
theme_economist() + xlim(-100,100) + theme(axis.title.x=element_blank(),
axis.text.y=element_blank(), axis.ticks.y=element_blank()) + ylab(temp_billtitle)
}
```

Examples:

```
> display_dots(34,39,"Covid",2)
> display_dots(47,57,"GOP Healthcare Bill",1)
```

Visualising the averages, cleaning the data and graph

The article mainly specifies more than 70% support for the Covid Bill. In order to create a matching graph, we have created the averages table that contains “Average Support”, “Average Oppose” and “Average Margin”. Individual results are grouped by Bill and average is calculated as below.

Averages of different polling firms (%) (install dplyr first)

1. `new_table<-econdata%>%select(Bill,support,oppose,margin)` - Create a table with only Bill,support,oppose and margin
   From the main data, select only Bill, Support Oppose and Margin columns and create a new table. For “%>%”, we installed the “dplyr” library.
2. Define mean_support - `mean_support<-aggregate(x=new_table$support,by=list(new_table$Bill),FUN=mean)`
   This aggregate function groups the results by the bill and takes average. For “Average Support”, “Average Oppose” and “Average Margin”, the function has to be called separately.
3. Define also others, change only support with oppose or margin
4. Create a new table for averages, assign first average support to that table so that following column could be added next to it - `means_table<-mean_support`
5. Combine first table with other averages - `merge(mean_support,mean_oppose,mean_margin,by="Bill")`
   After 3 different tables are created, we merge them by the “Bill” since it's the common column among these tables.
6. Change column names - `colnames(mean_support)<-c("Bill","Average Support")`
7. Sort the table according to Average Margin
   To display a better result, we sorted the result according to the “Average Margin”

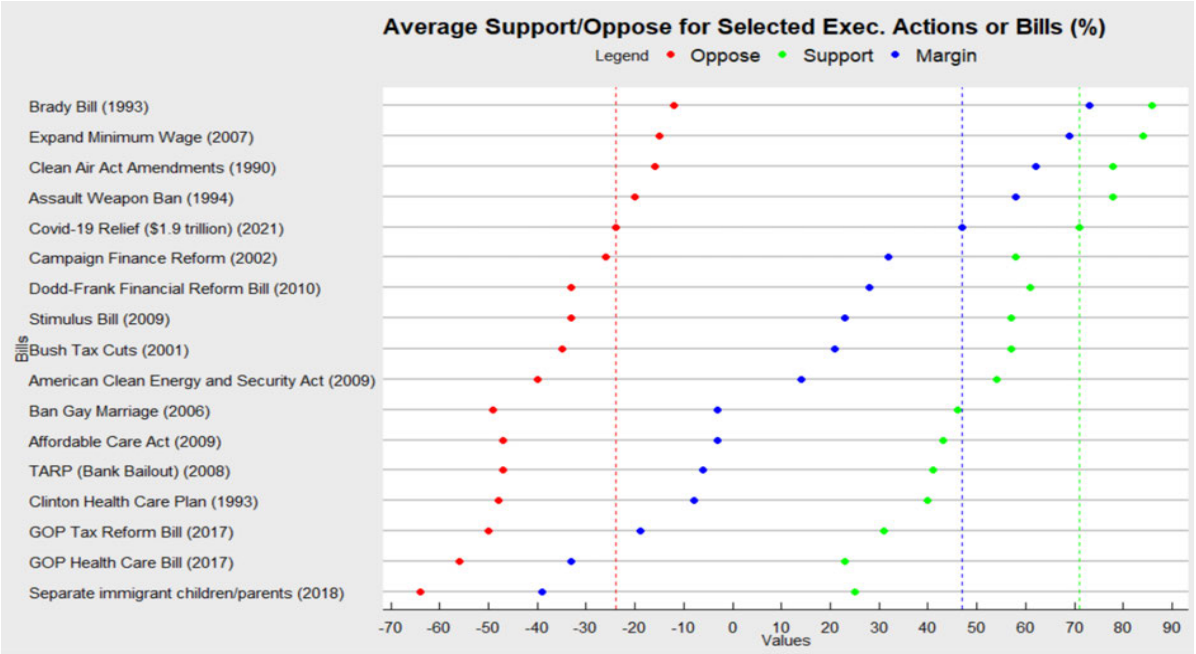
Bill	Average Support	Average Oppose	Average Margin
Separate immigrant children/parents (2018)	25	64	-39.2
GOP Health Care Bill (2017)	23	56	-32.8
GOP Tax Reform Bill (2017)	31	50	-19.0
Clinton Health Care Plan (1993)	40	48	-8.2
TARP (Bank Bailout) (2008)	41	47	-5.7
Affordable Care Act (2009)	43	47	-3.3
Ban Gay Marriage (2006)	46	49	-3.0
American Clean Energy and Security Act (2009)	54	40	14.3
Bush Tax Cuts (2001)	57	35	21.4
Stimulus Bill (2009)	57	33	23.3
Dodd-Frank Financial Reform Bill (2010)	61	33	27.8
Campaign Finance Reform (2002)	58	26	31.5
Covid-19 Relief ($1.9 trillion) (2021)	71	24	47.0
Assault Weapon Ban (1994)	78	20	58.0
Clean Air Act Amendments (1990)	78	16	62.0
Expand Minimum Wage (2007)	84	14	69.0
Brady Bill (1993)	86	12	73.3

The main ggplot code

```
ggplot(aver_table, aes(y=aver_table$Bill))[1]
+geom_point(aes(x=aver_table$`Average Margin`, color="red"), size=2)
+geom_point(aes(x=aver_table$`Average Support`, colour ="green"), size=2)
+geom_point(aes(x=aver_table$`Average Oppose`, colour ='blue'), size=2)[2]
+labs(x="Values",y="Bills",colour="Legend")[3]
+scale_x_continuous(breaks = seq(-100, 100, by = 10))[4]
+ggtitle("Average Support/Oppose for Selected Exec. Actions or Bills (%)")[5]
+geom_vline(xintercept=71,col="green",linetype=2)
+geom_vline(xintercept=-24,col="red",linetype=2)
+geom_vline(xintercept=47,col="blue",linetype=2)5+theme_economist_white()[6]
+scale_colour_manual(labels = c("Oppose", "Support","Margin"), values = c("red",
"green","blue"))[7]
```

Instead of creating layers, the graph is created with one code including all layers

1. From the table, “Bill” names are listed on y axis with ggplot as basis.
2. On the x axis, “Average Support”, “Average Oppose” and “Average Margin” are included with geom_point.
3. Labels for the axis and Legend are added.
4. In order to display the x axis with intervals by 10, scale_x_continuous is used.
5. ggtitle for the heading of the graph
6. To specify Covid Bill average results, we added vertical lines on the graph with geom_vline
7. Theme_economist is available under ggthemes packages. This code turns the graph into an economist style graph.
8. Finally, names of the color legend are adjusted.
9. For the blue format, the same path is followed but theme_economist is used instead of theme_economist_white.



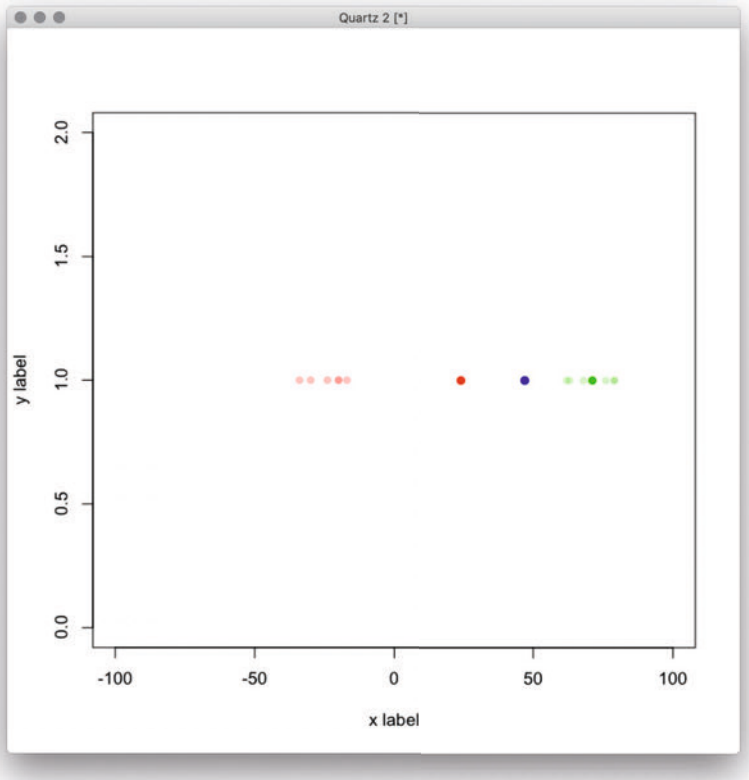
Plotting one bill with plot() and points()

This code generates a dotplot for one bill.

Requires:

- data frame named “polls” containing data from the CSV file from C. Warshaw

```
# create an empty plot
plot(NULL, xlim=c(-100,100), ylim=c(0,2), ylab="y label", xlab="x label")

# this example is for the COVID bill, note that the outer for loop (i) currently only runs once

for(i in 1:1) {

        temp_meansupport <- mean(polls$support[34:39])
        temp_meanoppose <- mean(polls$oppose[34:39])
        temp_meanmargin <- temp_meansupport - temp_meanoppose

        for(j in 34:39){
                points(polls$support[j],i,col=rgb(0,1,0,0.3),pch=16)
                points(-polls$oppose[j],i,col=rgb(1,0,0,0.3),pch=16)

        }

        points(temp_meansupport,i,col=rgb(0,1,0,1),pch=19)
        points(temp_meanoppose,i,col=rgb(1,0,0,1),pch=19)
        points(temp_meanmargin,i,col=rgb(0,0,1),pch=19)
}
```


Quartz 2 [*]
y label
2.0
1.5
1.0
0.5
0.0
-100
-50
0
50
100
x label

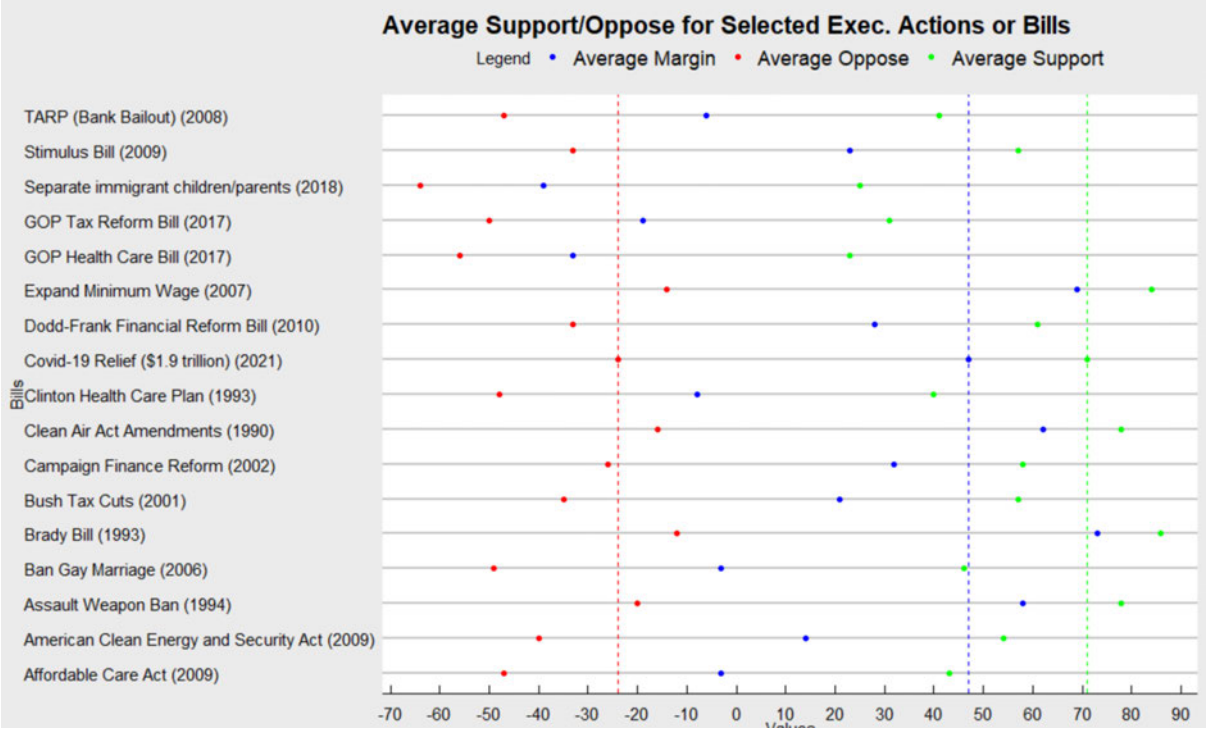
Results

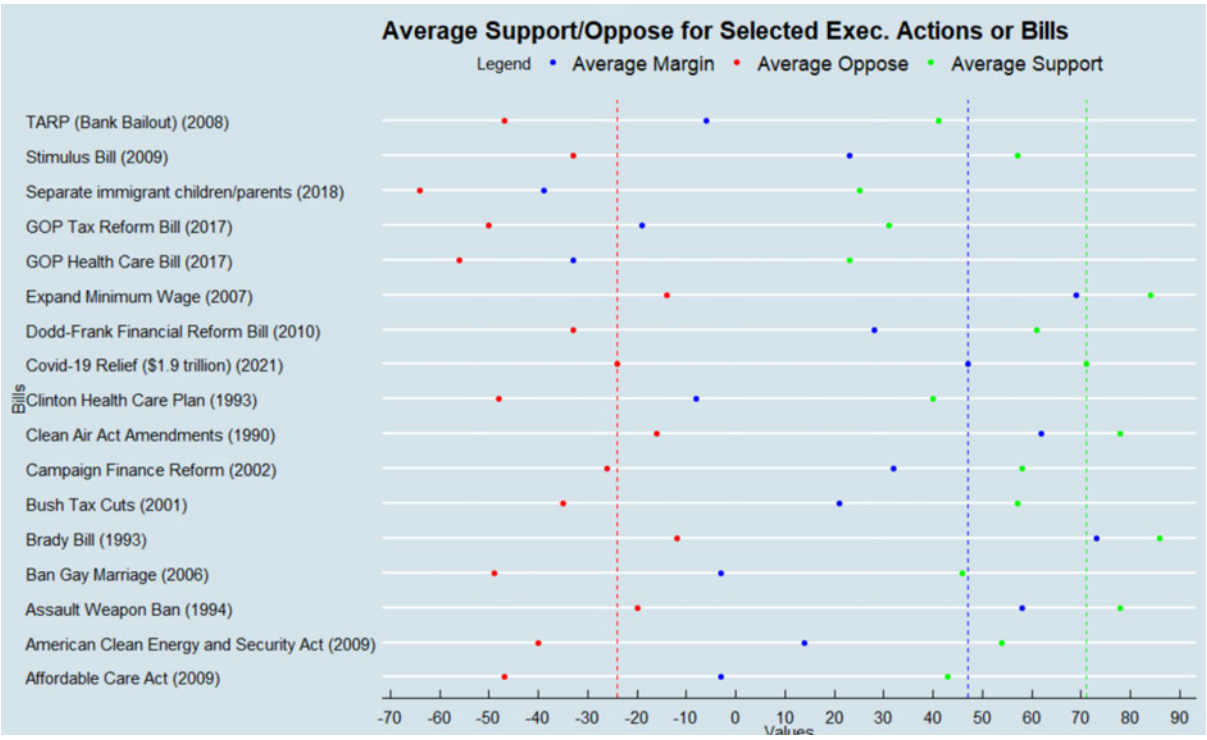
Plotting all bills with ggplot2

Averages of different polling firms (%) (install dplyr first)

1. aver_table<-econdata%>%select(Bill,support,oppose,margin) - Create a table with only Bill,support,oppose and margin
2. Define mean_support - mean_support<-aggregate(x=new_table$support,by=list(new_table$Bill),FUN=mean)
3. Define also others, change only support with oppose or margin
4. Create a new table for averages, assign first average support to that table so that following column could be added next to it - means_table<-mean_support
5. Combine first table with other averages - merge(mean_support,mean_oppose,mean_margin,by="Bill")
6. Change column names - colnames(mean_support)<-c("Bill","Average Support")
7. Sort the table according to Average Margin

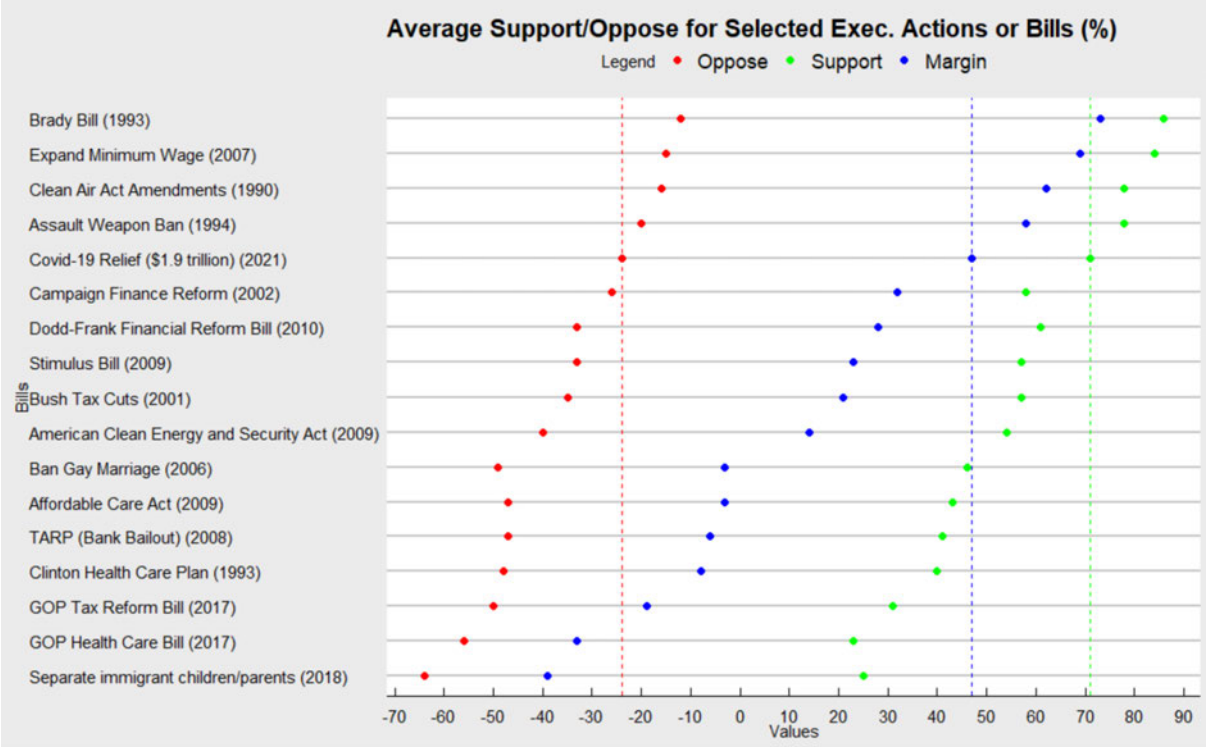
Bill	Average Support	Average Oppose	Average Margin
Separate immigrant children/parents (2018)	25	64	-39.2
GOP Health Care Bill (2017)	23	56	-32.8
GOP Tax Reform Bill (2017)	31	50	-19.0
Clinton Health Care Plan (1993)	40	48	-8.2
TARP (Bank Bailout) (2008)	41	47	-5.7
Affordable Care Act (2009)	43	47	-3.3
Ban Gay Marriage (2006)	46	49	-3.0
American Clean Energy and Security Act (2009)	54	40	14.3
Bush Tax Cuts (2001)	57	35	21.4
Stimulus Bill (2009)	57	33	23.3
Dodd-Frank Financial Reform Bill (2010)	61	33	27.8
Campaign Finance Reform (2002)	58	26	31.5
Covid-19 Relief ($1.9 trillion) (2021)	71	24	47.0
Assault Weapon Ban (1994)	78	20	58.0
Clean Air Act Amendments (1990)	78	16	62.0
Expand Minimum Wage (2007)	84	14	69.0
Brady Bill (1993)	86	12	73.3


Average Support/Oppose for Selected Exec. Actions or Bills
Legend
Average Margin
Average Oppose
Average Support
TARP (Bank Bailout) (2008)
Stimulus Bill (2009)
Separate immigrant children/parents (2018)
GOP Tax Reform Bill (2017)
GOP Health Care Bill (2017)
Expand Minimum Wage (2007)
Dodd-Frank Financial Reform Bill (2010)
Covid-19 Relief ($1.9 trillion) (2021)
Clinton Health Care Plan (1993)
Clean Air Act Amendments (1990)
Campaign Finance Reform (2002)
Bush Tax Cuts (2001)
Brady Bill (1993)
Ban Gay Marriage (2006)
Assault Weapon Ban (1994)
American Clean Energy and Security Act (2009)
Affordable Care Act (2009)
Bills
-70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90

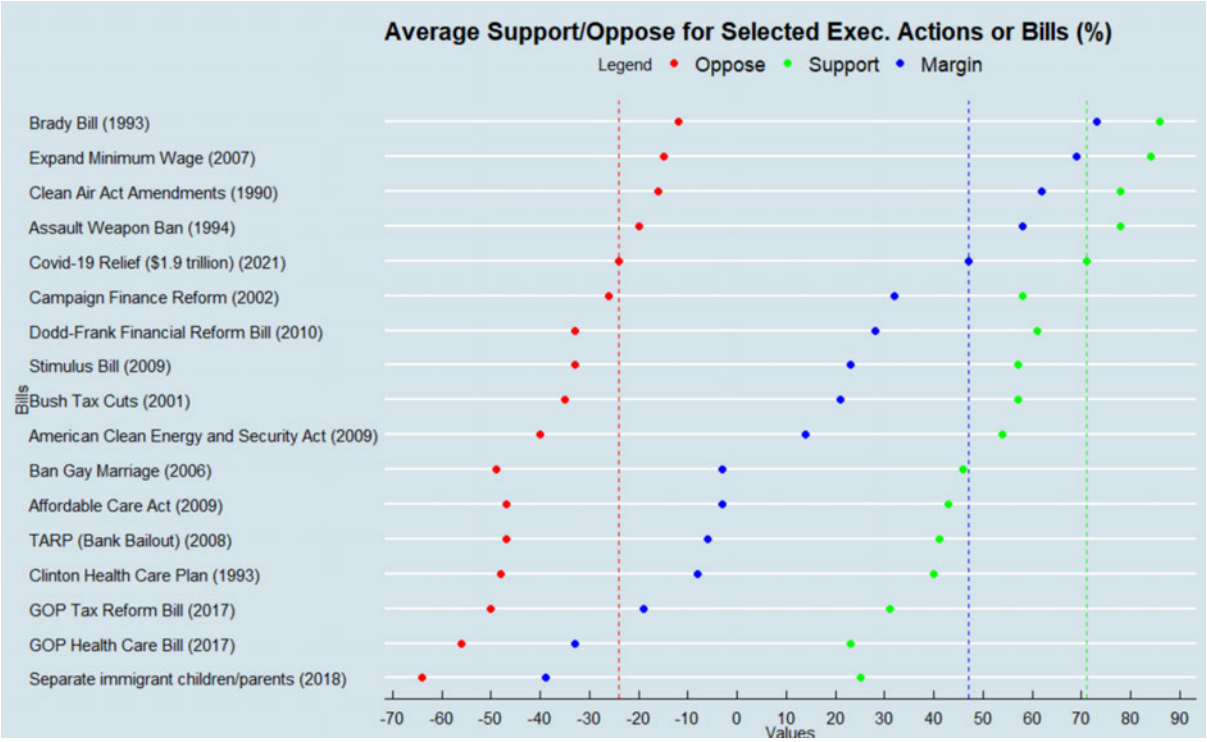

Average Support/Oppose for Selected Exec. Actions or Bills
Legend
Average Margin
Average Oppose
Average Support
TARP (Bank Bailout) (2008)
Stimulus Bill (2009)
Separate immigrant children/parents (2018)
GOP Tax Reform Bill (2017)
GOP Health Care Bill (2017)
Expand Minimum Wage (2007)
Dodd-Frank Financial Reform Bill (2010)
Covid-19 Relief ($1.9 trillion) (2021)
Clinton Health Care Plan (1993)
Clean Air Act Amendments (1990)
Campaign Finance Reform (2002)
Bush Tax Cuts (2001)
Brady Bill (1993)
Ban Gay Marriage (2006)
Assault Weapon Ban (1994)
American Clean Energy and Security Act (2009)
Affordable Care Act (2009)
Bills
-70 -60 -50 -40 -30 -20 -10 0 10 20 30 40 50 60 70 80 90

One big ggplot code (ggplot and ggthemes (for theme_economist) packages installed)

```
ggplot(aver_table, aes(y=aver_table$Bill))
+ geom_point(aes(x=aver_table$`Average Margin`, color="red"), size=2)
+geom_point(aes(x=aver_table$`Average Support`, colour ="green"), size=2)
+geom_point(aes(x=aver_table$`Average Oppose`, colour ='blue'), size=2)
+labs(x="Values",y="Bills",colour="Legend")
+scale_x_continuous(breaks = seq(-100, 100, by = 10))
+ggtitle("Average Support/Oppose for Selected Exec. Actions or Bills (%)")
+geom_vline(xintercept=71,col="green",linetype=2)
+geom_vline(xintercept=-24,col="red",linetype=2)
+geom_vline(xintercept=47,col="blue",linetype=2)
+theme_economist_white()
+scale_colour_manual(labels = c("Oppose", "Support","Margin"), values = c("red",
"green","blue"))
```



```
ggplot(aver_table, aes(y=aver_table$Bill))
+geom_point(aes(x=aver_table$`Average Margin`, color="red"), size=2)
+geom_point(aes(x=aver_table$`Average Support`, colour ="green"), size=2)
+geom_point(aes(x=aver_table$`Average Oppose`, colour ='blue'), size=2)
+labs(x="Values",y="Bills",colour="Legend")
+scale_x_continuous(breaks = seq(-100, 100, by = 10))
+ggtitle("Average Support/Oppose for Selected Exec. Actions or Bills (%)")
+geom_vline(xintercept=71,col="green",linetype=2)
+geom_vline(xintercept=-24,col="red",linetype=2)
+geom_vline(xintercept=47,col="blue",linetype=2)
+theme_economist()
+scale_colour_manual(labels = c("Oppose", "Support","Margin"), values = c("red",
"green","blue"))
```



Plotting all bills with plot() and points()

This code generates a dotplot for all bills.

Requires:

- data frame named “polls” containing data from the CSV file from C. Warshaw
- data frame named “bills_index” containing the start and end rows of each bill (see Google Drive > Datasets), sorted by average net support

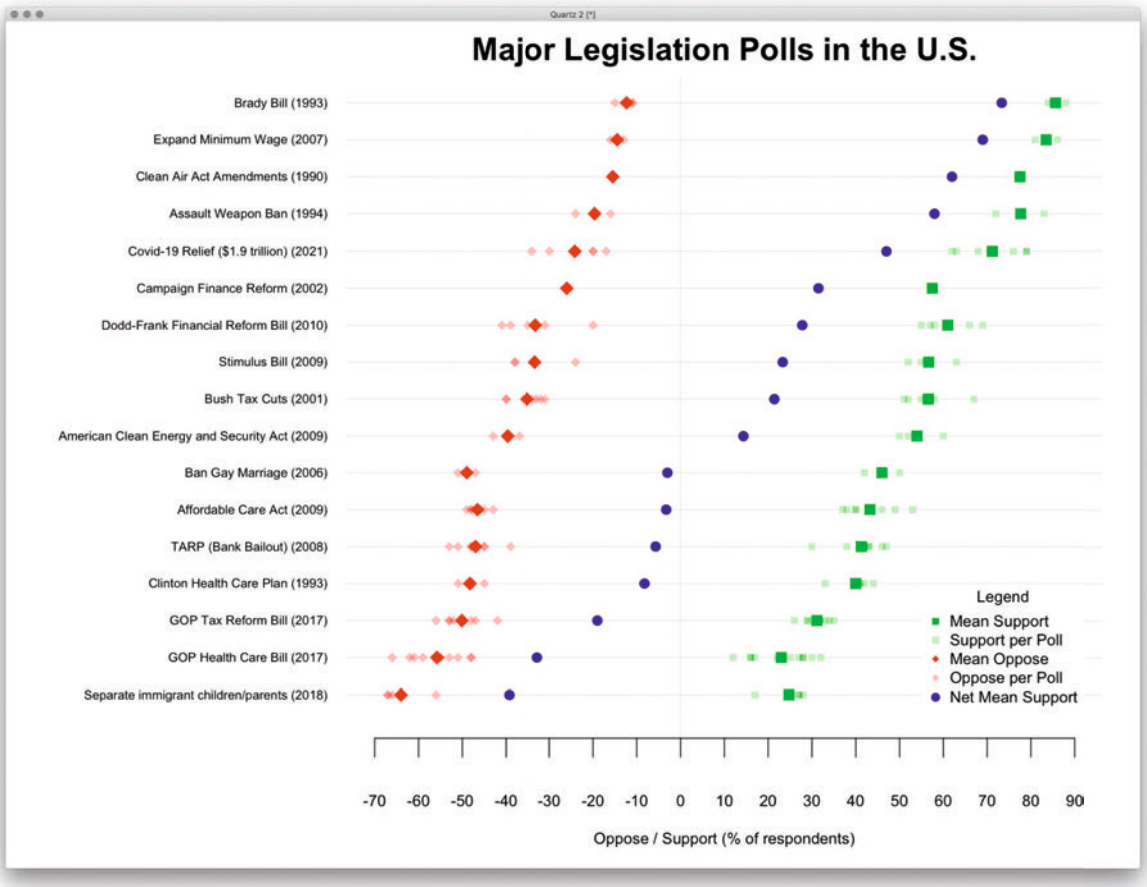
Note: read the csv for the bills_index with this line because it contains ; as separators:

```
bills_index <- read.csv("bills-index.csv",sep=";")
```

Code for the graph:

```
# Create a new quartz device (mac only) for more space and higher resolution of graph
quartz(width=6,height=4.5,dpi=300)

# Set margin (bottom,left,top,right)
par(mar = c(3, 9, 1.5, 1))

# Set up the plot
plot(NULL, xlim=c(-70,90), ylim=c(0,17), cex.lab=0.7, yaxt='n', xaxt='n', ylab="",
xlab="",axes=FALSE)
abline(v=0,col=rgb(0.9,0.9,0.9),lwd=0.5)
axis(side=2,at=c(1:17),cex.axis=0.4, las=2,
labels=factor(bills_index$Bill,bills_index$Bill),tick=FALSE,line=-0.5)
axis(side=1,cex.axis=0.5,at=seq(-70,90,by=10),lwd=0.5,line=-0.5)
title(xlab="Oppose / Support (% of respondents)", line=1.5, cex.lab=0.5)

# Draw the dots
# Outer for-loop (i): go through each line of the data frame containing the start and end rows
of each bill, already sorted by mean net support
# Inner for-loop (j): for each bill, go through each line of the data frame containing all polls to
draw individual polls

for(i in 1:17) {

        temp_meansupport <- mean(polls$support[bills_index$Start[i]:bills_index$End[i]])
        temp_meanoppose <- mean(polls$oppose[bills_index$Start[i]:bills_index$End[i]])
        temp_meanmargin <- temp_meansupport - temp_meanoppose

        abline(h=i,col=rgb(0.9,0.9,0.9),lwd=0.5)

        for(j in bills_index$Start[i]:bills_index$End[i]) {

                points(polls$support[j],i,col=rgb(0,0.8,0,0.3),pch=15,cex=0.5)
                points(-polls$oppose[j],i,col=rgb(1,0,0,0.3),pch=18,cex=0.7)
```

```
        }

        points(temp_meansupport,i,col=rgb(0,0.8,0,1),pch=15,cex=0.8)
        points(-temp_meanoppose,i,col=rgb(1,0,0,1),pch=18,cex=1.0)
        points(temp_meanmargin,i,col=rgb(0,0,1),pch=19,cex=0.6)

}



# Add legend and title

legend(55,4,legend=c("Mean Support","Support per Poll","Mean Oppose", "Oppose per
Poll", "Mean Net Support"), col=c(rgb(0,0.8,0,1), rgb(0,0.8,0,0.3),rgb(1,0,0,1), rgb(1,0,0,0.3),
rgb(0,0,1)), pch=c(15,15,18,18,19), cex=0.5, bg="white",box.lty=0,title="Legend")

title(main="Major Legislation Polls in the U.S.", cex.main=1)
```



Discussion

As stated in the introduction part, the purpose of the project is to verify whether 70% of the population supports Biden's COVID-19 relief bill with the help of a clear and persuasive visualization. Throughout this EDA project, the project team strives to improve the visualization of the Economist article by using the same dataset as the journalist from the Economist.

Although one could debate whether the final two visualizations, created by the team, are indeed much better than the original one, we as the team believe that they are an improvement. The process of understanding the background and trying various solutions out enabled us to figure out what is the most suitable graph and the most efficient logic in presenting visualization. The two improved visualizations will create better support of the claim of the original article since the 70% support of the COVID-19 relief bill can be indicated in both graphs. The added legend and improved color spectrum, are in our point of view, make it less confusing for the readers than the original visualization does.

Not only the graph but also our own data analysis and visualization skills are improved as well. We are more familiar with the R packages and what they are able to do. We achieved this not only through listening to the lectures and doing Datacamp assignments, but also through trying out different commands to create the result we prefer more, and seeking help from the mass online communities when we feel trapped due to the lack of knowledge.

Based on our own experience and the group discussion, we have made several initial remarks comparing the graphs created by ggplot() and base R plot().

There are several advantages of the R plot() that we identified. The first one is, as already mentioned in the class, plot() is more intuitive to use than ggplot(). We have experienced the same and therefore conclude that plot() can be understood by novices who encounter R or coding the very first time. The second advantage, and also the primary reason for us to consider using plot() in our project, is the fact that by using plot(), elements can be added to the plot after the plot is being drawn. More layers can be added, and it makes the use of for-loops possible.

One clear advantage of ggplot() is the rich themes one can install when creating plots. The choices of theme enabled us to make the graphs look great with less code. Another advantage of ggplot() is, it can be used in a leaner way, which suggests it requires less coding, as long as one knows what the correct codes that one needs are.

In our project, to come to the final results, we created separate tables consisting of necessary data for our final visualizations. However, creating extra tables could be caused by our limited knowledge in using R. During our discussion, the team believes that ggplot() does not necessarily require a supplementary table to achieve the final result. With accurate coding, one could include the sorting for the table in the aggregate function. On the other hand, this might be harder to achieve when using plot(). The for-loop can complicate the process since different bills have a different number of polls.

Limitations

Several limitations were identified throughout the project. The original graph from the article gives us a set ground for us to develop our improved visualization. On the one hand, the original graph provides us ideas on how we could construct our visualizations; on the other hand, it limited us to picture outside the box and create more different designs. The improved graphs are highly linked to the original article since we want to verify the 70% support of the COVID-19 relief Bill stated in the article. The constraints of developing our own findings based on the existing ones limited us to be more creative in terms of trying new plots or new logics.

The dataset that we received from Prof. Warshaw is a relatively small and straightforward one. This gives us less opportunity to explore the various ways to understand it. Our focus on this dataset then focuses more on the process since we already have a draft vision of how the final result should look at the start of the project. The challenge becomes how to document and present the process clearly when it is completed.

Another limitation regarding the original dataset is that we have full access and can use the data; however, we cannot validate where the data come from and don't know the complete source of the data.

Our beginner-level R knowledge also hinders us on the way to recreate a persuasive visualization. It occurs to us often that we have an idea of the effect we want on the graph but don't know precisely how we get there. We are confident that we could use it better in telling comprehensive stories once we get more familiar with R.

What are our limitations?

- Graph needs to be linked to the article
- Small dataset – project focuses on process
- We have access to the data but no capability to validate it (no access to full source)
- We are at a beginner level of R

Conclusion

Overall, we set out to improve the piece of data journalism found in The Economist. We believe that we achieved an improvement in terms of supporting the main point of the article. We obtained the original dataset, critiqued the original visualization, and explored different ways of visualizing the data in the context of the article.

We created two final graphs, both of which are dotplots which plot all bills. The first version was created using ggplot() and contains each bill's average margin, average support, and average oppose values. This plot already supports the claim that roughly 70% of the public support the COVID-19 relief bill of 2021. The second version was created using plot() (base R) and contains not only each bill's averages but also each individual poll's support and oppose values. It was created using a for-loop.

Finally, we discussed the different approaches to visualizing the data. We found that there are different advantages and disadvantages to using ggplot() vs. plot(). We are more informed now as to when to use which approach.

We are particularly proud of obtaining the real dataset and quickly prototyping different graphs in an experimental approach. We are also satisfied with the team process, having updated each other regularly.

References

Arnold, J. B., Daroczi, G., Werth, B., Weitzner, B., Kunst, J., Auguie, B., Rudis, B., Wickham, H., Talbot, J., & London, J. (2021). ggthemes: Extra Themes, Scales and Geoms for “ggplot2.” https://cran.r-project.org/package=ggthemes

Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. In T. Azzam & S. Evergreen (Eds.), Data visualization, part 1. New Directions for Evaluation (Issue 139, pp. 7–32). https://doi.org/10.1002/ev

Carpendale, M. (2003). Considering Visual Variables as a Basis for Information Visualisation. https://doi.org/10.11575/PRISM/30495

Cornwell, S., & Brice, M. (2021, March 10). Biden’s $1.9 trillion COVID-19 bill wins final approval in House. Reuters. https://www.reuters.com/article/us-health-coronavirus-usa-congress-idUSKBN2B215E

Dörk, M., Feng, P., Collins, C., & Carpendale, S. (2013). Critical InfoVis. CHI ’13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA ’13, 2189. https://doi.org/10.1145/2468356.2468739

Penumaka, E. (2021). The American Rescue Plan is Popular and Holds Bipartisan Support. Data for Progress. https://www.dataforprogress.org/blog/2021/3/9/voters-support-american-rescue-plan-25pjp

President Biden (@POTUS). (2021). Tweet “From launching community vaccination centers around the country to deploying ….” Twitter.Com. https://twitter.com/POTUS/status/1369426527893528576

Team, R. C. (2021). R: A language and environment for statistical computing. https://www.r-project.org/

The Economist. (2021, February). Joe Biden’s $1.9trn stimulus package is one of the most popular bills in decades. The Economist. https://www.economist.com/graphic-detail/2021/02/26/joe-bidens-19trn-stimulus-package-is-one-of-the-most-popular-bills-in-decades

Tufte, E. (1990). Envisioning Information. Graphics Press LLC.

Warshaw, C. (2021). Dataset: Major Legislation Polls (U.S.).

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A Grammar of Data Manipulation. R package version 1.0.6.

Appendices

Appendix 1

Original dataset (see major_legislation_polls.csv as a separate file)

Appendix 2

Index table used for plotting with plot() and for-loop (see bills-index.csv as a separate file)