



R Workshop: Visualizing FB Misinformation

3/4/2022

ETHITECH

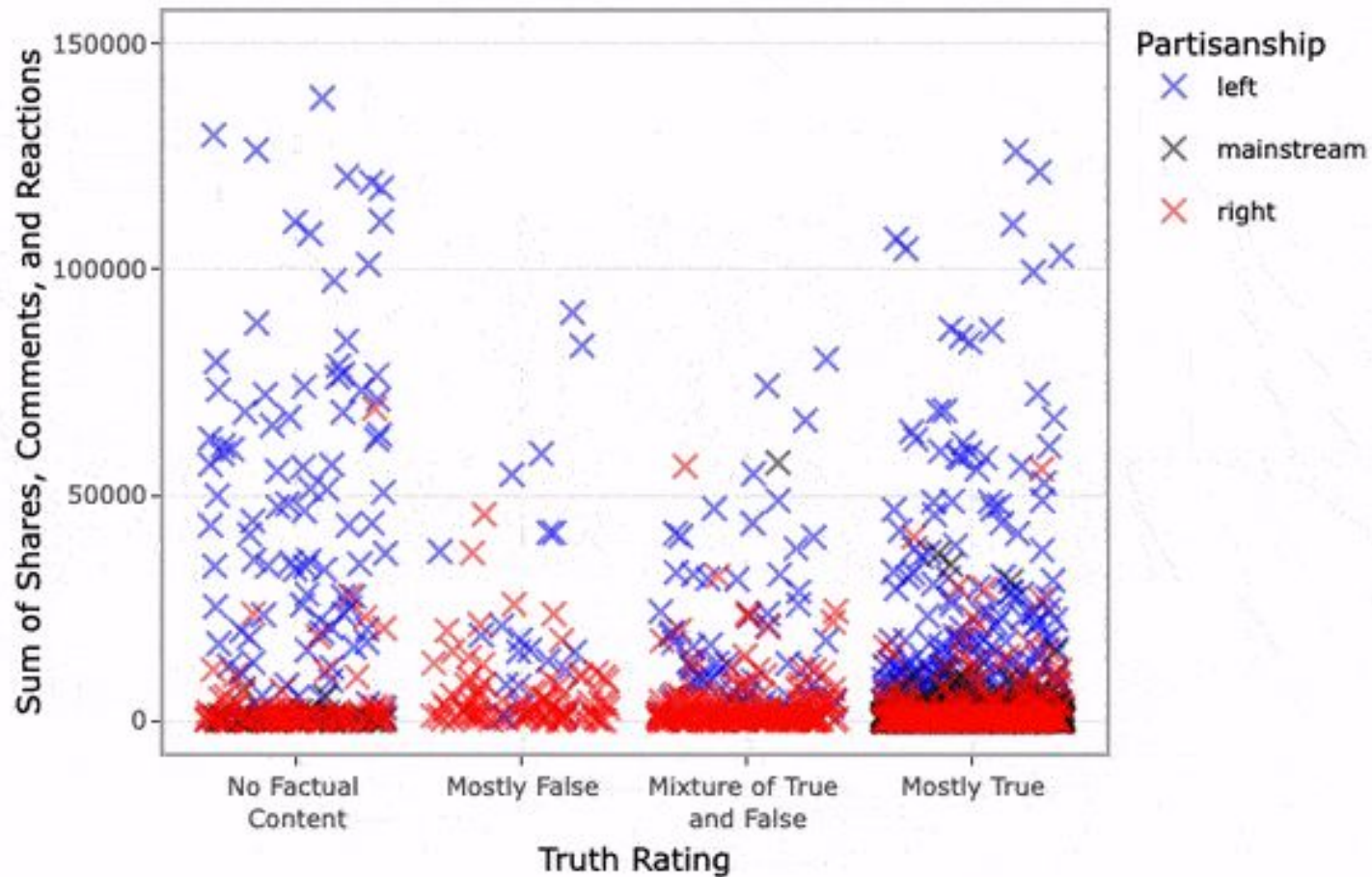
GO

AGENDA



1. Overview of project
 - a. What types of misinformation gain the most traction?
2. Searching for data
3. Coding in RStudio
4. Publishing your work
5. Further inspiration

The Spread of Facebook Misinformation



SEARCHING FOR DATA

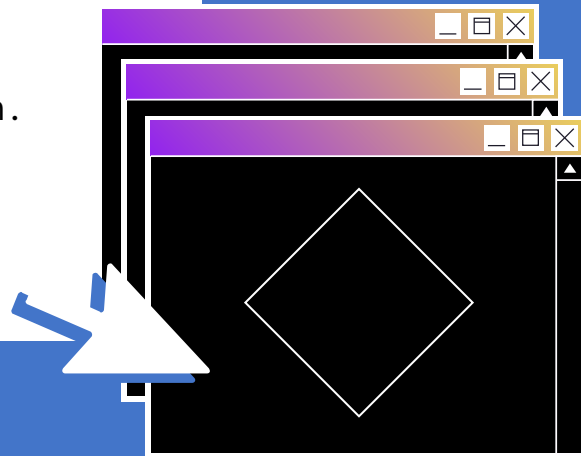
Start with sites dedicated to datasets:

- datasetsearch.research.google.com/
- kaggle.com/datasets
- public.opendatasoft.com/explore/

Use keywords and filters to narrow your search.

Today, we'll be working with

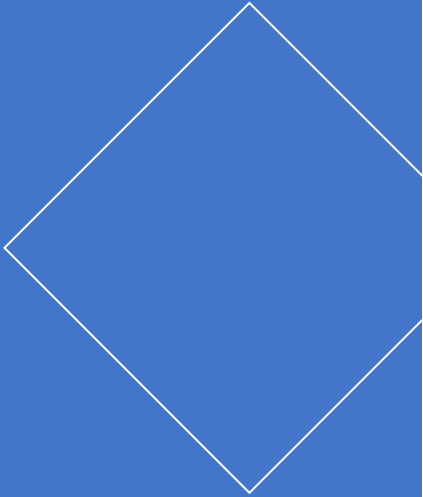
kaggle.com/mrisdal/fact-checking-facebook-politics-pages



- rstudio.cloud
(Must create an account)
- github.com/SamanthaLLee/facebook-misinformation
- rpubs.com
(Must create an account)

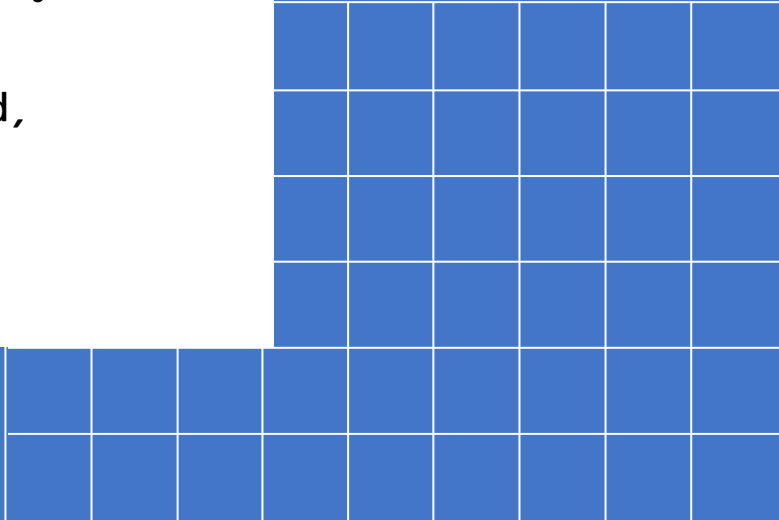
LINKS FOR TODAY ' S WORKSHOP

USING PACKAGES



```
If first time using,  
install.packages("ggplot2")  
install.packages("plotly")
```

```
If already installed,  
library(ggplot2)  
library(plotly)
```



LOADING DATA



```
# Load data
```

```
fbdata <- read.csv("https://raw.githubusercontent.com/SamanthaLLee/  
facebook-misinformation/main/data/facebook-fact-check.csv")
```

```
summary(fbdata)
```

```
View(fbdata)
```

account_id	post_id	Category	Page	Post.URL	Date.Published	Post.Type
Min. :6.232e+10	Min. :5.511e+14	Length:2282	Length:2282	Length:2282	Length:2282	Length:2282
1st Qu.:1.145e+14	1st Qu.:1.247e+15	Class :character	Class :character	Class :character	Class :character	Class :character
Median :1.841e+14	Median :1.291e+15	Mode :character	Mode :character	Mode :character	Mode :character	Mode :character
Mean :1.867e+14	Mean :3.300e+15					
3rd Qu.:3.469e+14	3rd Qu.:1.541e+15					
Max. :4.401e+14	Max. :1.015e+16					

Rating	Debate	share_count	reaction_count	comment_count	activity_count	Rating_num
Length:2282	Length:2282	Min. : 1	Min. : 2.0	Min. : 0.0	Min. : 9.0	Min. :0.000
Class :character	Class :character	1st Qu.: 24	1st Qu.: 149.0	1st Qu.: 37.0	1st Qu.: 256.8	1st Qu.:2.000
Mode :character	Mode :character	Median : 96	Median : 545.5	Median : 131.5	Median : 881.0	Median :3.000
		Mean : 4045	Mean : 5364.3	Mean : 516.1	Mean : 9975.2	Mean :2.454
		3rd Qu.: 739	3rd Qu.: 2416.8	3rd Qu.: 390.2	3rd Qu.: 3855.5	3rd Qu.:3.000
		Max. :1088995	Max. :456458.0	Max. :159047.0	Max. :1704500.0	Max. :3.000
		NA's :70	NA's :2	NA's :2	NA's :70	

	account_id	post_id	Category	Page	Post.URL	Date.Published	Post.Type	Rating
1	1.840966e+14	1.035058e+15	mainstream	ABC News Politics	https://www.facebook.com/ABCNewsPolitics/posts/1...	2016-09-19	video	no factual content
2	1.840966e+14	1.035269e+15	mainstream	ABC News Politics	https://www.facebook.com/ABCNewsPolitics/posts/1...	2016-09-19	link	mostly true
3	1.840966e+14	1.035306e+15	mainstream	ABC News Politics	https://www.facebook.com/ABCNewsPolitics/posts/1...	2016-09-19	link	mostly true

C&P: <https://raw.githubusercontent.com/SamanthaLLee/facebook-misinformation/main/data/facebook-fact-check.csv>

MEASURING ACTIVITY

- Sum up shares, reactions, and comments and create a new column in the data frame to store the overall activity count

```
# Manipulate data
fbdata$activity_count = fbdata$share_count +
  fbdata$reaction_count + fbdata$comment_count
```

share_count	reaction_count	comment_count	activity_count
1	33	34	68
34	63	27	124
35	170	86	291
568	3188	2815	6571
23	28	21	72
46	409	105	560
7	62	64	133
7	39	6	52

OPERATIONALIZING TRUTH

- Map truth ratings to numbers

```
fbdata$Rating_num = ifelse(fbdata$Rating == "mostly true", 3,  
                           ifelse(fbdata$Rating == "mixture of true and false", 2,  
                                   ifelse(fbdata$Rating == "mostly false", 1, 0)))
```

SUBSETTING DATA

- Examine specific categories by creating new data frames

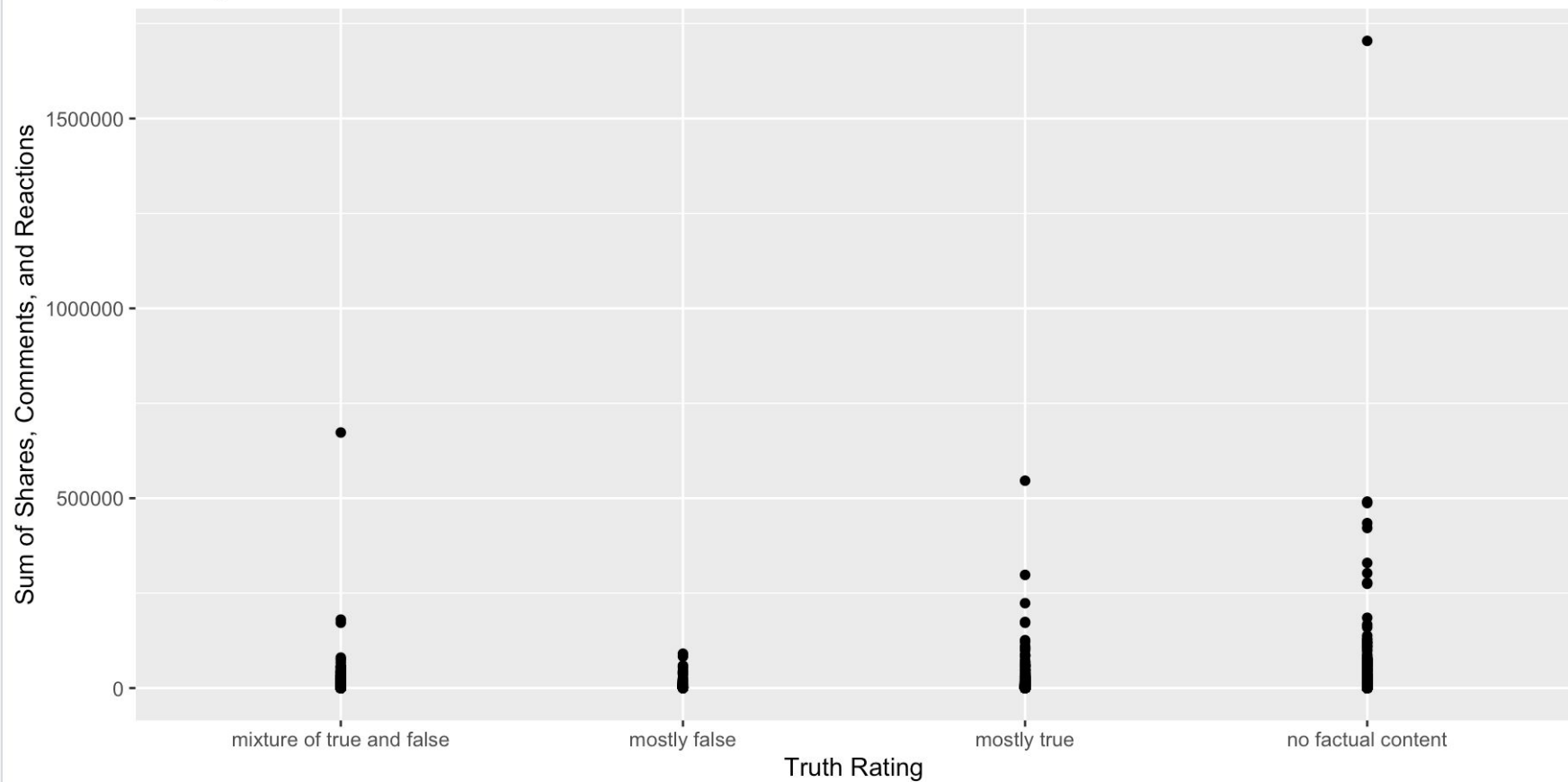
```
# Subset into categories of interest  
nonpartisan <- subset(fbdata, Category == "mainstream")  
liberal <- subset(fbdata, Category == "left")  
conservative <- subset(fbdata, Category == "right")
```

PLOTTING DATA



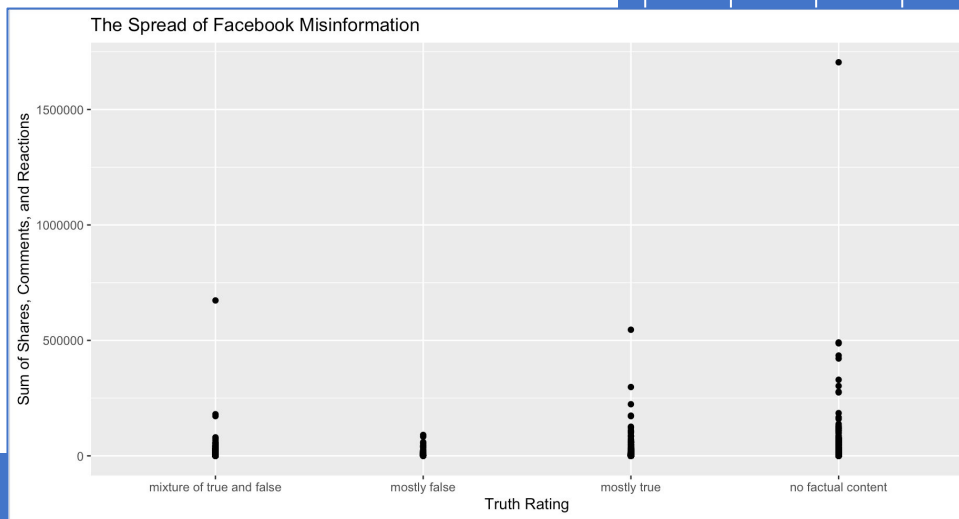
```
# Plot data
p <- ggplot(data=fbdata, mapping = aes(x=Rating, y=activity_count)) +
  geom_point(na.rm=T)+
  ggtitle("The Spread of Facebook Misinformation")+
  ylab("Sum of Shares, Comments, and Reactions")+
  xlab("Truth Rating")
p
```

The Spread of Facebook Misinformation



PLOT PROBLEMS

- Most y values are >500k
- Dots are stacked
- Truth rating in wrong order
- Boring

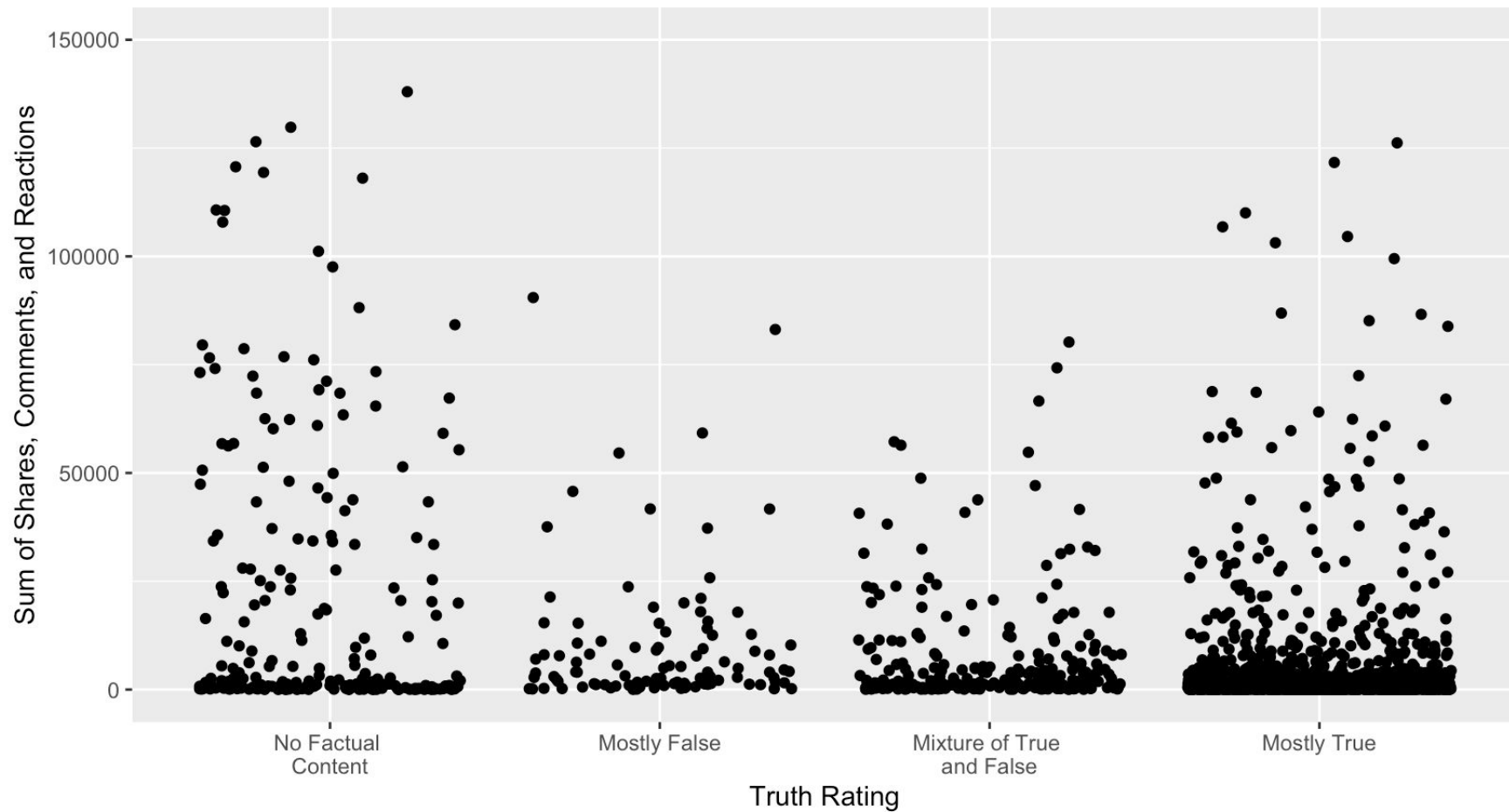


PLOTTING DATA, REVISED



```
# Plot data
p <- ggplot(data=fbdata, mapping = aes(x=Rating, y=activity_count)) +
  geom_point(na.rm=T, position = "jitter")+
  scale_x_discrete(limits = c("no factual content", "mostly false",
                              "mixture of true and false", "mostly true"),
                  labels = c("No Factual \nContent", "Mostly False",
                              "Mixture of True \nand False", "Mostly True"))+
  ggtitle("The Spread of Facebook Misinformation")+
  ylab("Sum of Shares, Comments, and Reactions")+
  xlab("Truth Rating")+
  ylim(0, 150000)
p
```

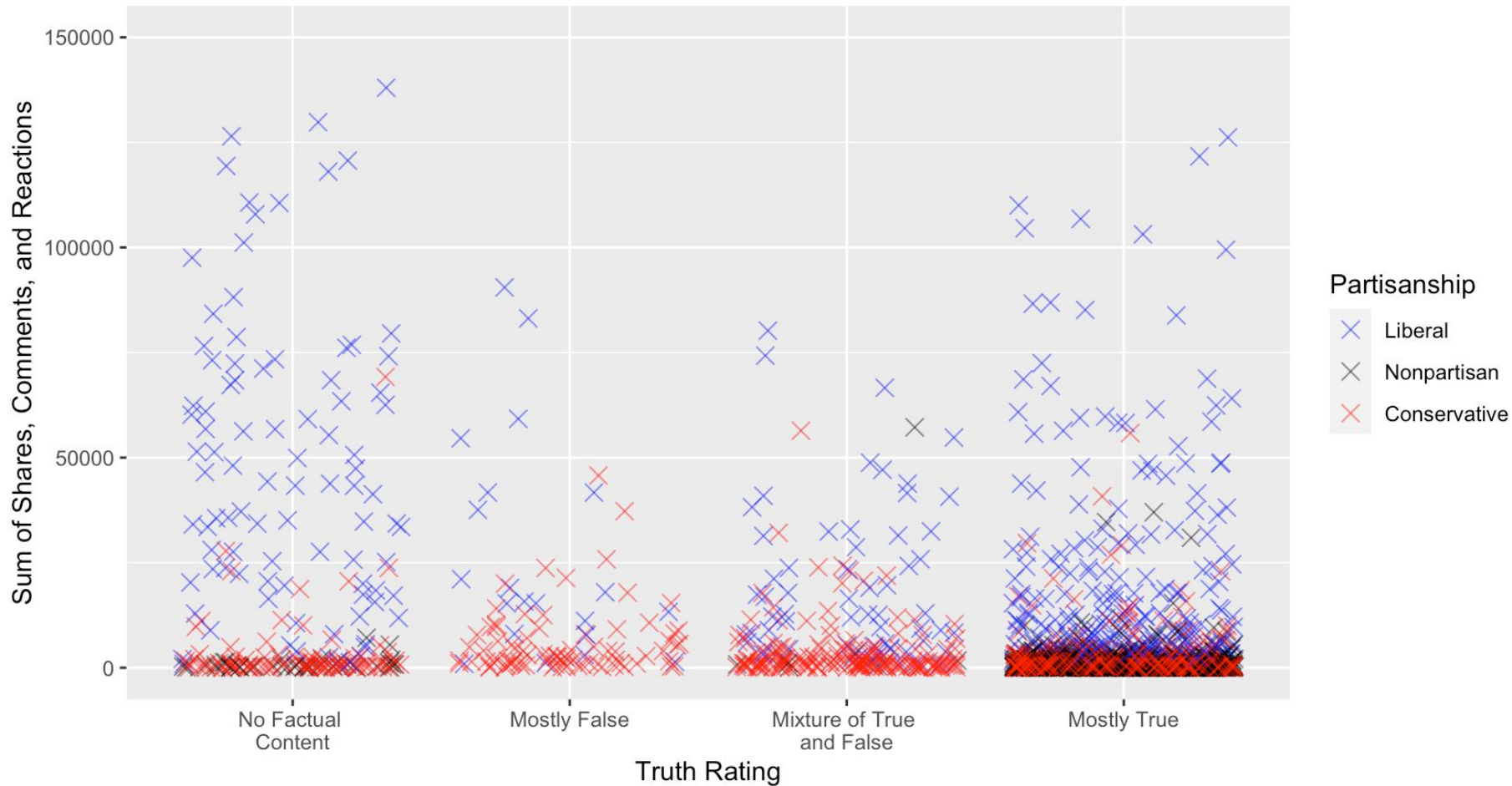
The Spread of Facebook Misinformation



PLOTTING DATA, TOUCHING UP

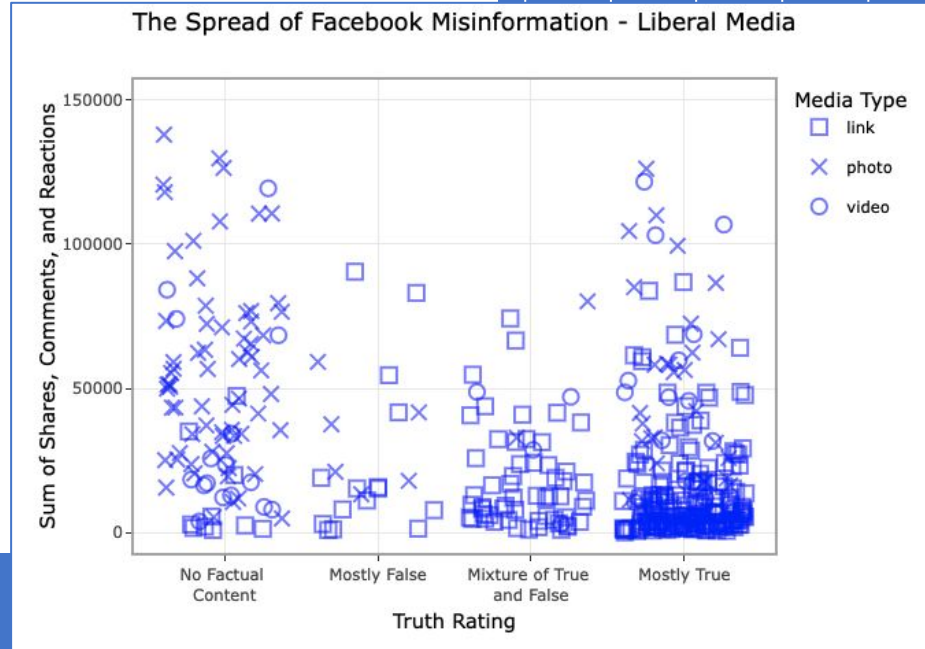
```
# Plot data
p <- ggplot(data=fbdata, mapping = aes(x=Rating, y=activity_count, color=Category)) +
  geom_point(na.rm=T, size=3, shape=4, alpha = .5, position = "jitter")+
  scale_x_discrete(limits = c("no factual content", "mostly false",
                              "mixture of true and false", "mostly true"),
                  labels = c("No Factual \nContent", "Mostly False",
                              "Mixture of True \nand False", "Mostly True"))+
  scale_color_manual(name = "Partisanship",
                    values=c('blue', 'black', 'red'),
                    limits = c("left", "mainstream", "right"),
                    labels = c("Liberal", "Nonpartisan", "Conservative"))+
  ggtitle("The Spread of Facebook Misinformation")+
  ylab("Sum of Shares, Comments, and Reactions")+
  xlab("Truth Rating")+
  ylim(0, 150000)
p
```


The Spread of Facebook Misinformation



LET 'S PLOT EVEN MORE

- Using our liberal, nonpartisan, and conservative subsets, we can examine each political category in its own plot.



PLOTTING LIBERAL DATA

change for conservative &
nonpartisan data frames accordingly

```
# Plot liberal data
l <- ggplot(data=liberal, mapping = aes(x=Rating, y=activity_count, shape=Post.Type)) +
  geom_point(na.rm=T, size=3, alpha = .5, color="blue", position = "jitter")+
  scale_x_discrete(limits = c("no factual content", "mostly false", "mixture of true and false", "mostly true"),
    labels = c("No Factual \nContent", "Mostly False", "Mixture of True \nand False", "Mostly True"))+
  ggtitle("The Spread of Facebook Misinformation - Liberal Media")+
  ylab("Sum of Shares, Comments, and Reactions")+
  xlab("Truth Rating")+
  ylim(0, 150000)
l
ggplotly(l)
```

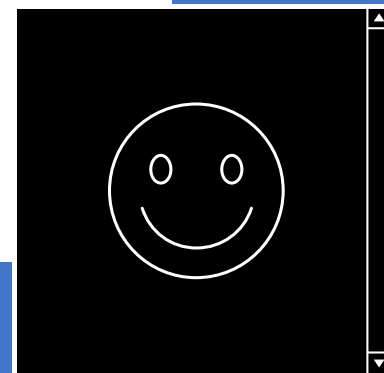
The Spread of Facebook Misinformation - Liberal Media



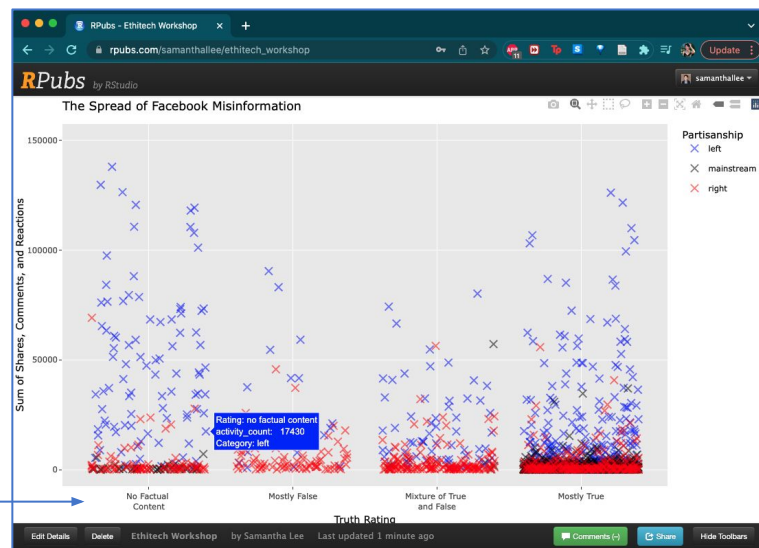
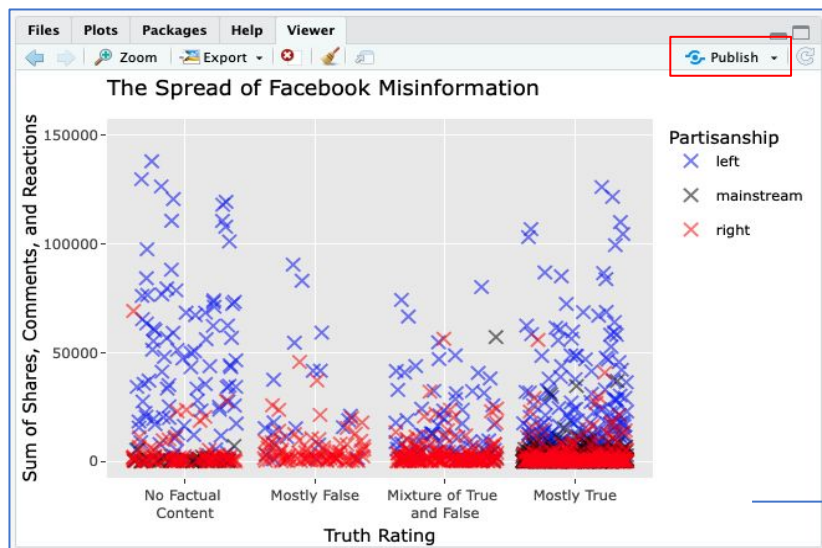


WHAT ' S NEXT?

You can directly publish your plots to <https://rpubs.com/> to share your work.



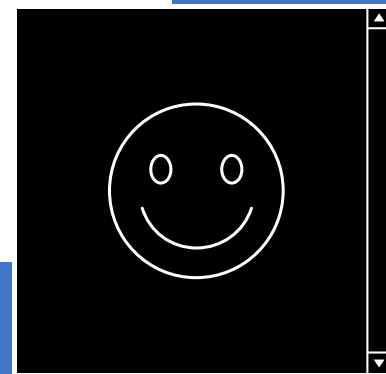
PUBLISHING DIRECTLY





R MARKDOWN

To have more control over what your R Pubs page displays, you can create an RMD file to better format, provide context, arrange data and code, and more.



Facebook Misinformation - Ethitech Workshop

Sam Lee

3/4/2022

Background

Misinformation has the power to create contention, influence public opinion, undermine democracy, and uplift malicious or ignorant parties. Its recent momentum has given reason to investigate the quality and correlations of news on popular sites like Facebook.

```
fbdata$activity_count = fbdata$share_count + fbdata$reaction_count + fbdata$comment_count  
summary(fbdata)
```

```
##      account_id      post_id      Category      Page  
## Min.   :6.232e+10  Min.   :5.511e+14  Length:2282  Length:2282  
## 1st Qu.:1.145e+14  1st Qu.:1.247e+15  Class :character  Class :character  
## Median :1.841e+14  Median :1.291e+15  Mode  :character  Mode  :character  
## Mean   :1.867e+14  Mean   :3.300e+15  
## 3rd Qu.:3.469e+14  3rd Qu.:1.541e+15  
## Max.   :4.401e+14  Max.   :1.015e+16  
##  
##      Post.URL      Date.Published      Post.Type      Rating  
## Length:2282      Length:2282      Length:2282      Length:2282  
## Class :character  Class :character  Class :character  Class :character  
## Mode  :character  Mode  :character  Mode  :character  Mode  :character  
##  
##  
##  
##  
##      Debate      share_count      reaction_count      comment_count  
## Length:2282      Min.   :0.0000000      Min.   :0.0000000      Min.   :0.0000000
```

[Edit Details](#)[Delete](#)

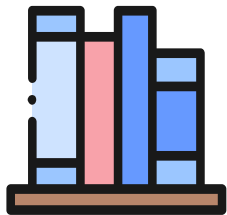
Facebook Misinformation - Ethitech Workshop

by Samantha Lee

last updated less than a minute ago

[Comments \(0\)](#)[Share](#)[Hide Toolbars](#)

FURTHER INSPIRATION



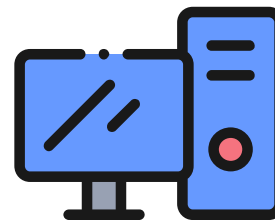
The R Graph Gallery

<https://www.r-graph-gallery.com/>



r/dataisbeautiful

<https://www.reddit.com/r/dataisbeautiful/>



Information is Beautiful

<https://informationisbeautiful.net/>



THANKS!

Any questions?

Find all code & slides @
[https://github.com/SamanthaLLee/
facebook-misinformation/workshop](https://github.com/SamanthaLLee/facebook-misinformation/workshop)

CREDITS: This presentation template was created by Slidesgo,
including icons by Flaticon, infographics & images by Freepik