

Lecture 8 Notes

2024-02-08

Downloading data

```
require(tidyverse)
```

```
## Loading required package: tidyverse
```

```
## Warning: package 'tidyverse' was built under R version 4.3.2
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.2      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2     3.4.4      ✓ tibble     3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
## ✓ purrr       1.0.1
```

```
## — Conflicts — tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## ⓘ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to
   become errors
```

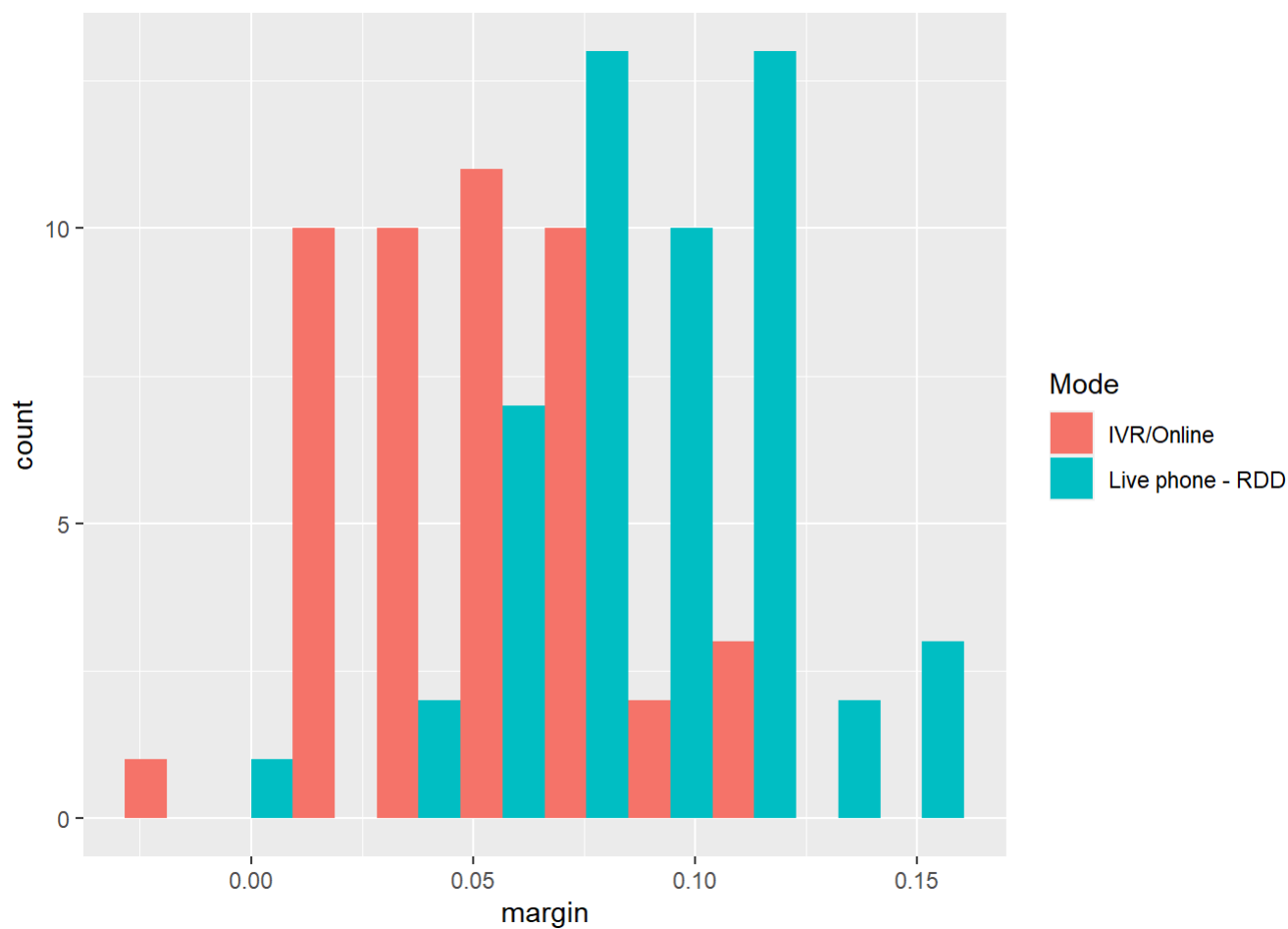
```
poll <- read_rds("https://github.com/jbisbee1/DS1000_S2024/raw/main/data/Pres2020_PV.Rds")

view(poll)

# Quick wrangle
poll <- poll %>%
  mutate(Trump = Trump/100,
         Biden = Biden/100,
         margin = Biden - Trump)
```

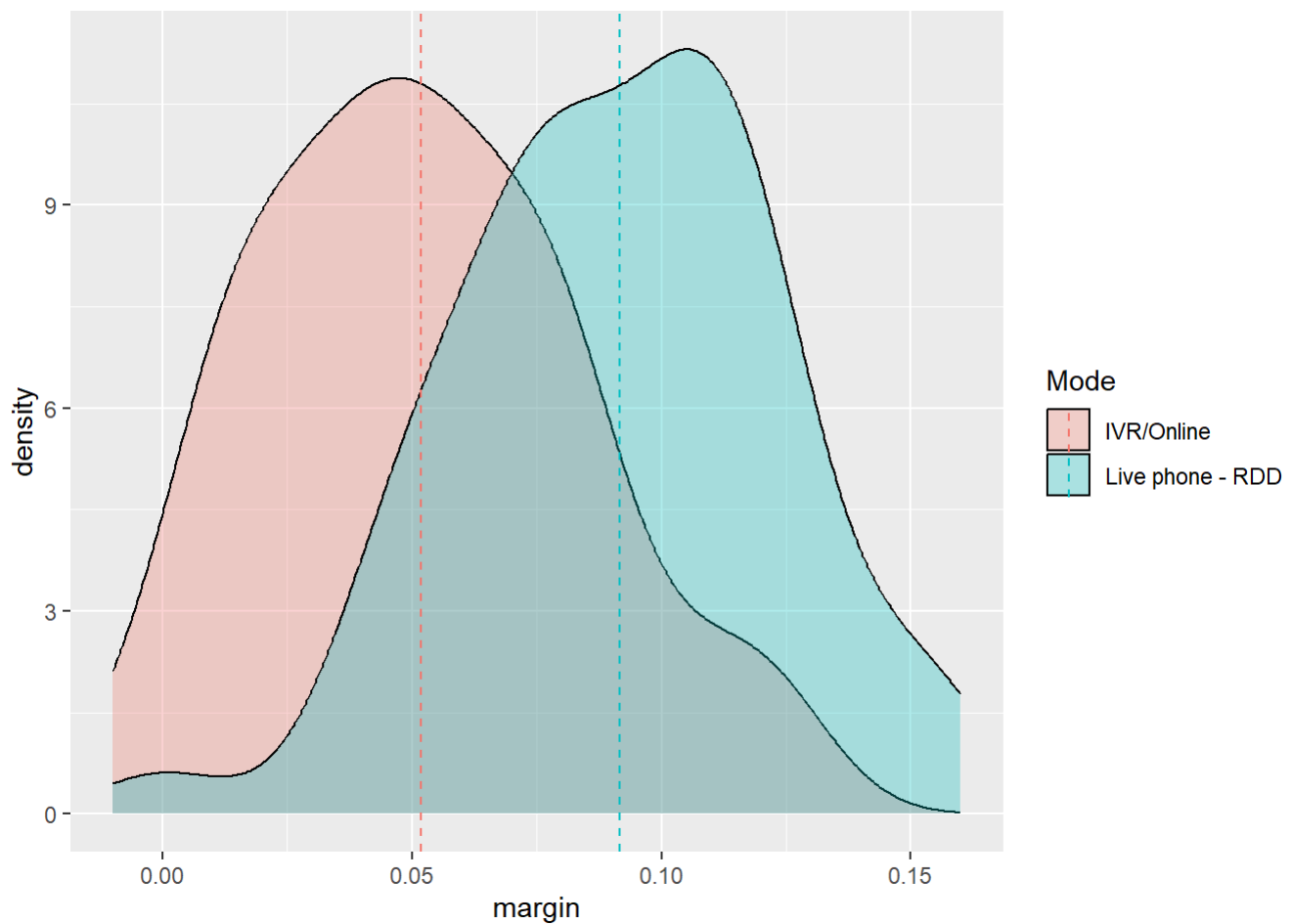
Multivariate Visualization: categorical by continuous

```
poll %>%
  filter(Mode == "IVR/Online" | Mode == "Live phone - RDD") %>%
  ggplot(aes(x= margin, fill = Mode)) +
  geom_histogram(bins = 10, position = 'dodge')
```



```
# Try a density instead
toplot <- poll %>%
  filter(Mode == "IVR/Online" | Mode == "Live phone - RDD")

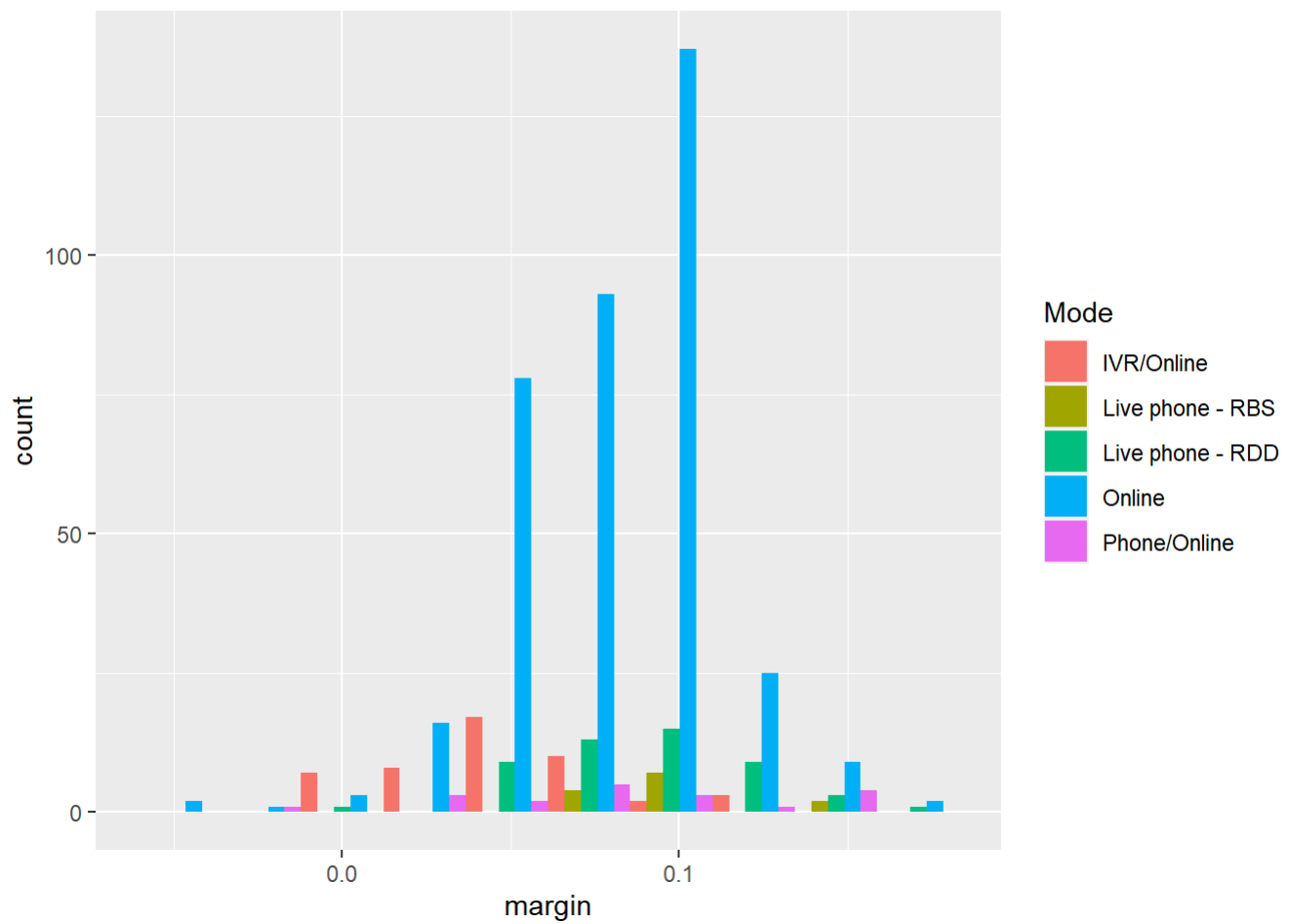
toplot %>%
  ggplot(aes(x= margin, fill = Mode)) +
  geom_density(alpha = .3) +
  geom_vline(data = toplot %>%
    group_by(Mode) %>%
    summarise(margin_avg = mean(margin, na.rm=T)),
    aes(xintercept = margin_avg, color = Mode),
    linetype = 'dashed')
```



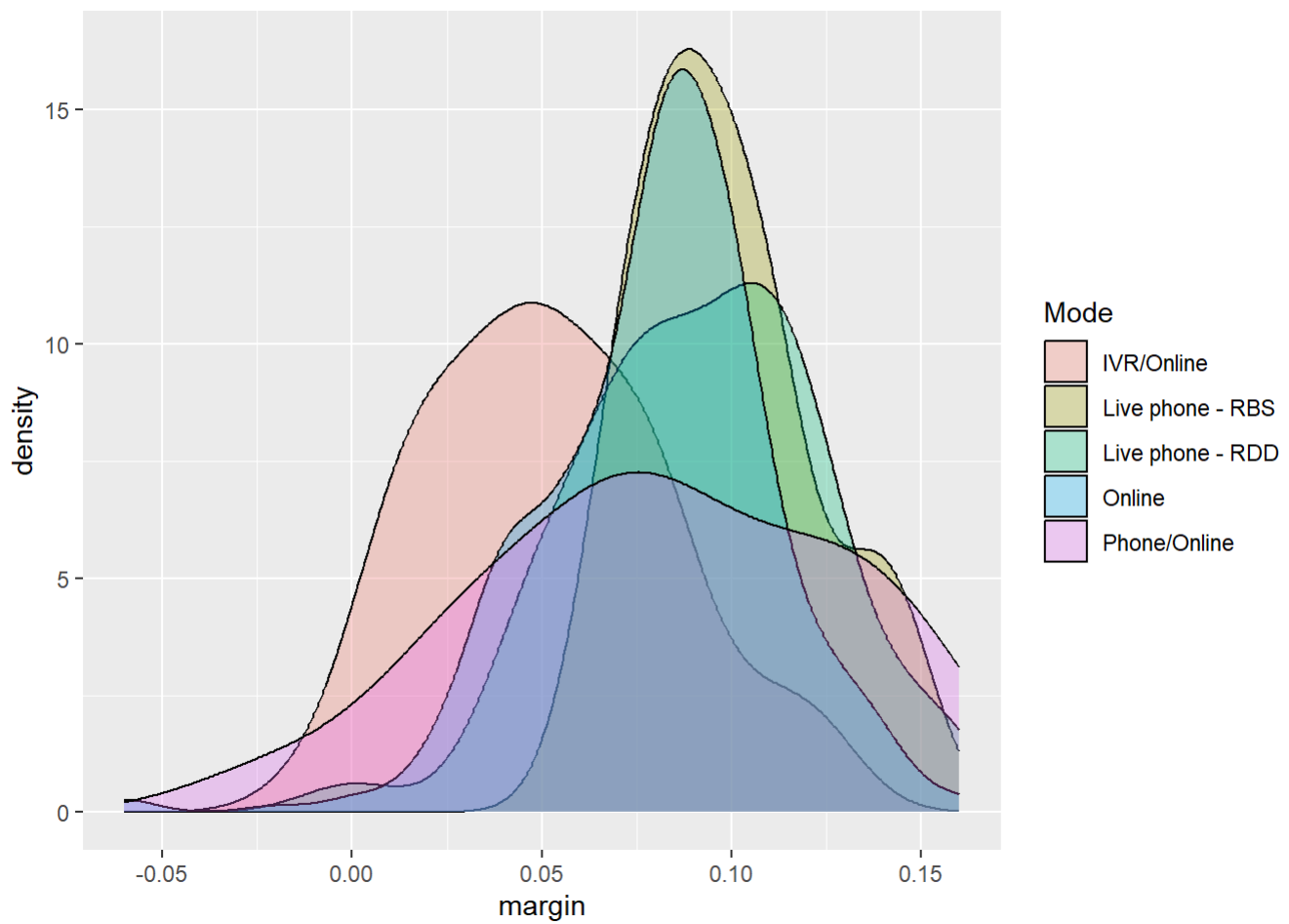
More categories

```
toKeep <- poll %>%
  count(Mode) %>%
  filter(n > 5) %>%
  filter(!is.na(Mode))

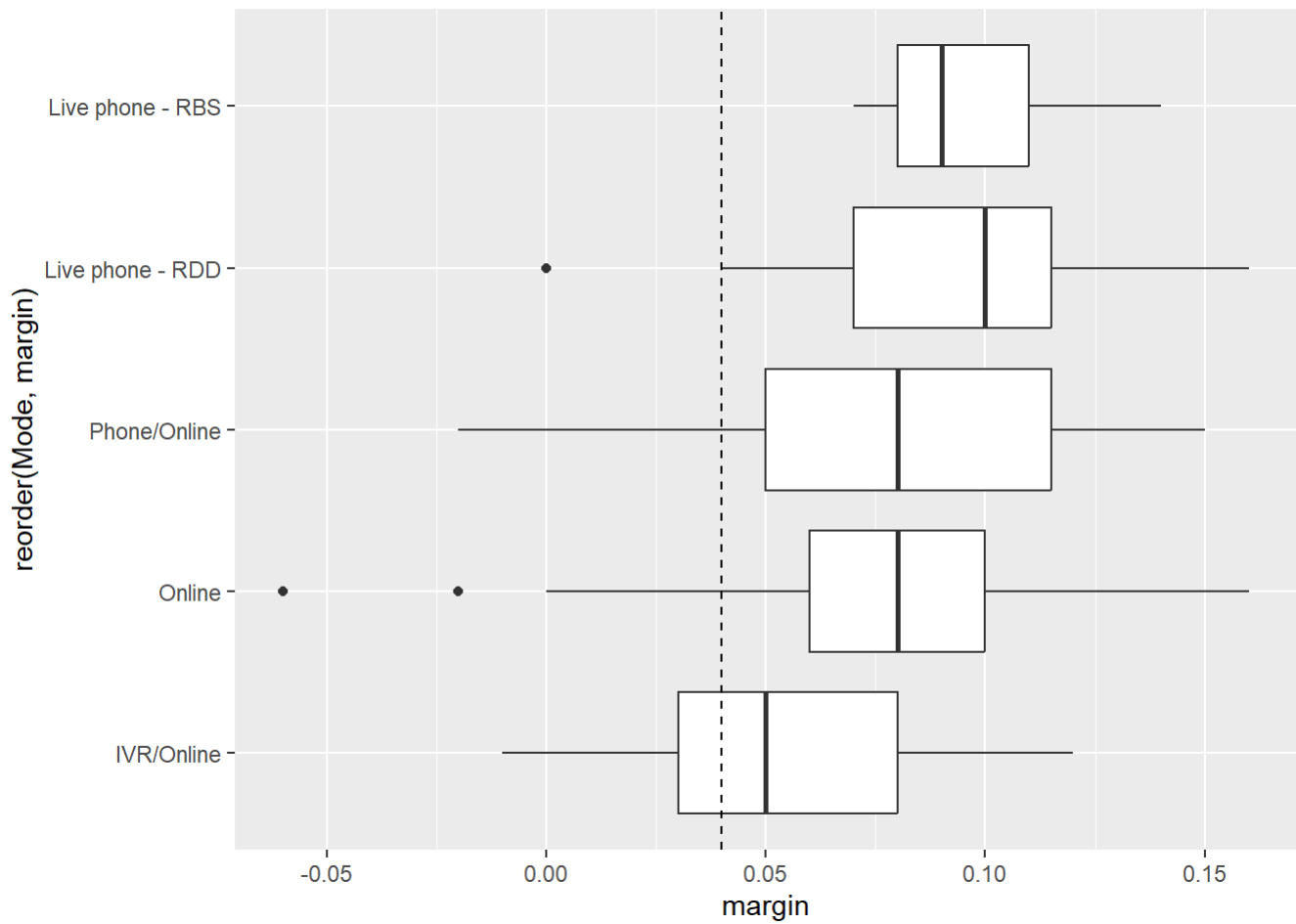
poll %>%
  filter(Mode %in% toKeep$Mode) %>%
  ggplot(aes(x = margin, fill = Mode)) +
  geom_histogram(bins = 10, position = 'dodge')
```



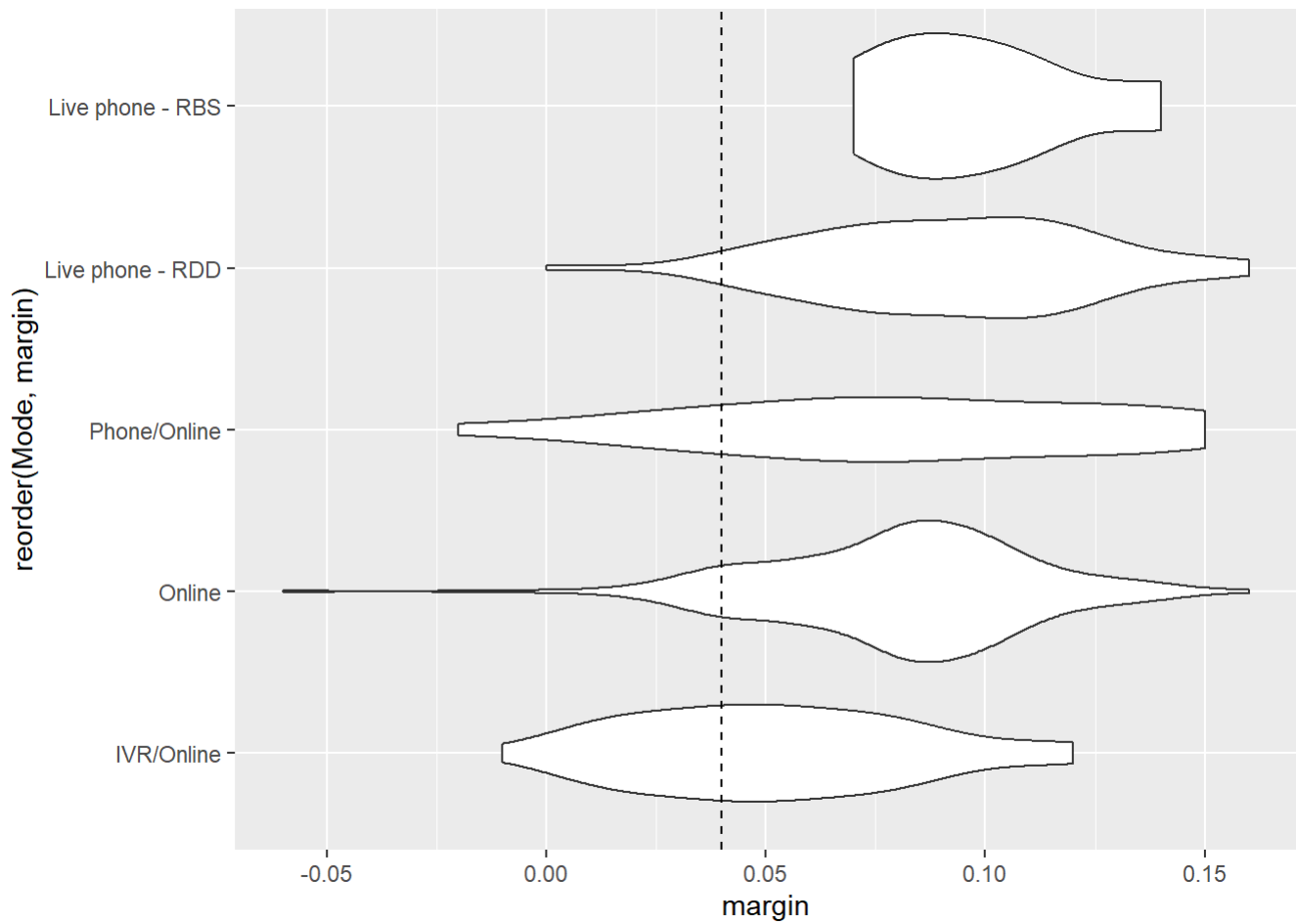
```
poll %>%
  filter(Mode %in% toKeep$Mode) %>%
  ggplot(aes(x = margin, fill = Mode)) +
  geom_density(alpha = .3)
```



```
# Better solution: geom_boxplot()
poll %>%
  filter(Mode %in% toKeep$Mode) %>%
  ggplot(aes(x = margin, y = reorder(Mode, margin))) +
  geom_boxplot() +
  geom_vline(xintercept = .04, linetype = 'dashed')
```

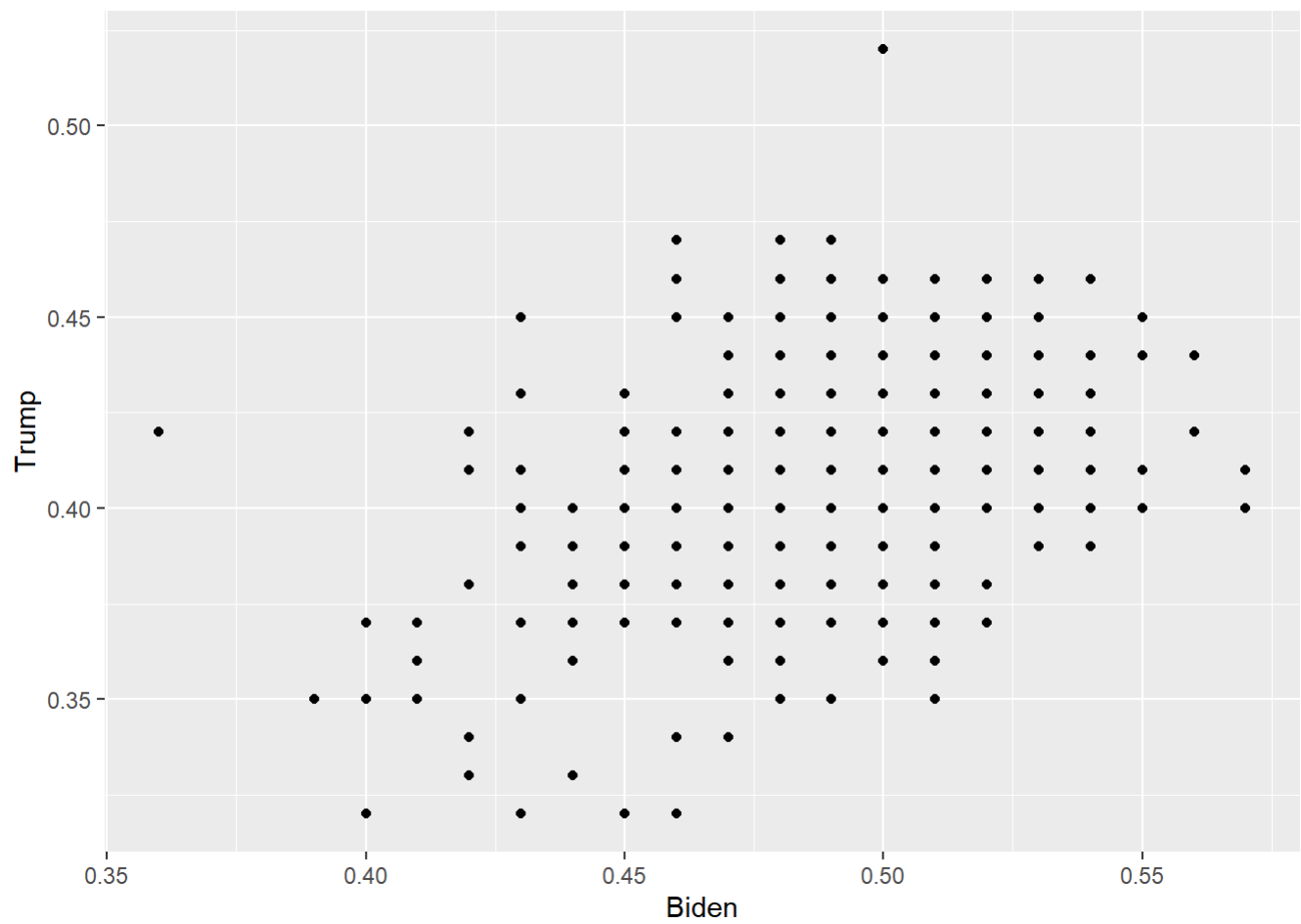


```
# Alternative: geom_violin()
poll %>%
  filter(Mode %in% toKeep$Mode) %>%
  ggplot(aes(x = margin, y = reorder(Mode, margin))) +
  geom_violin() +
  geom_vline(xintercept = .04, linetype = 'dashed')
```

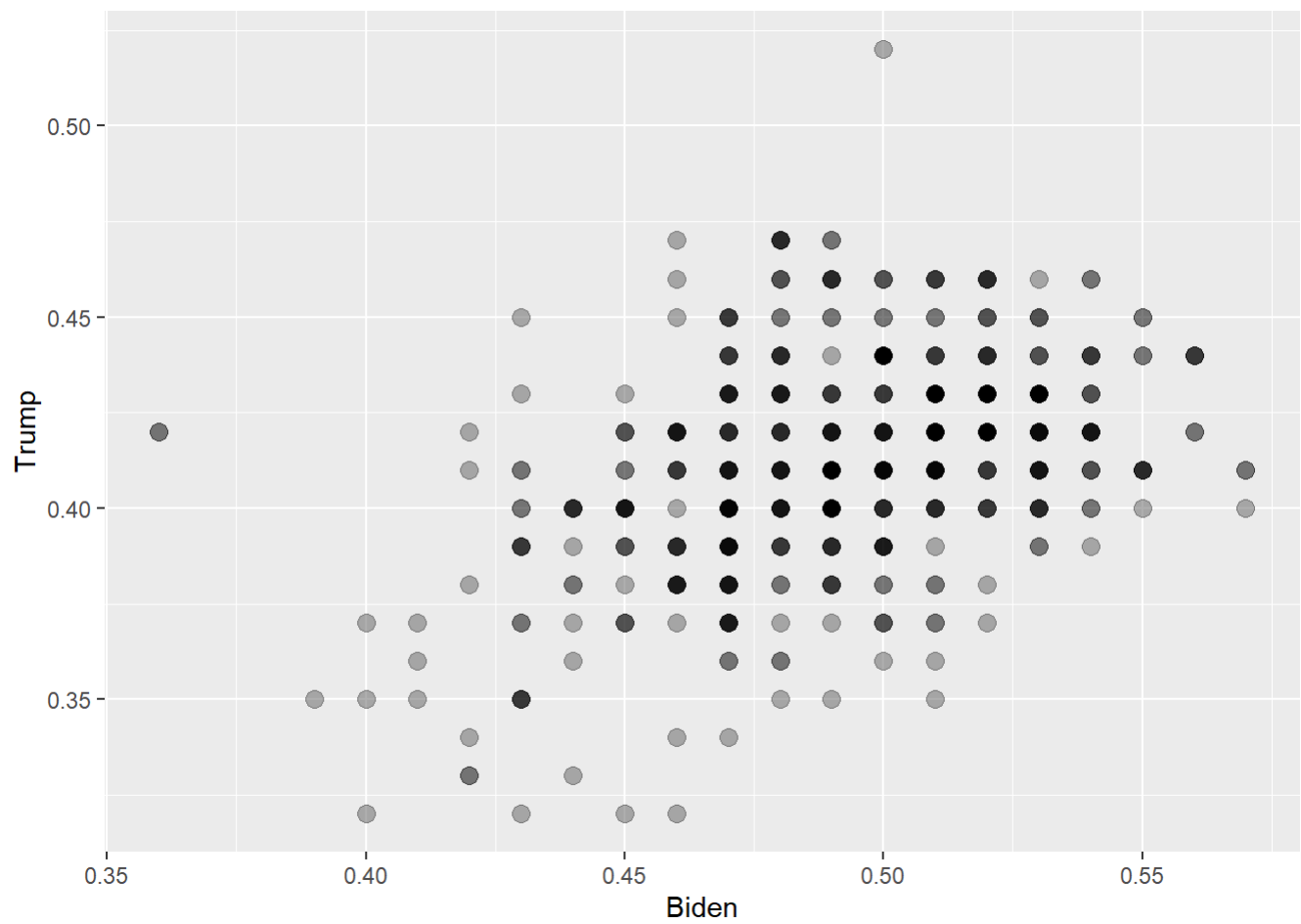


Politically biased polling

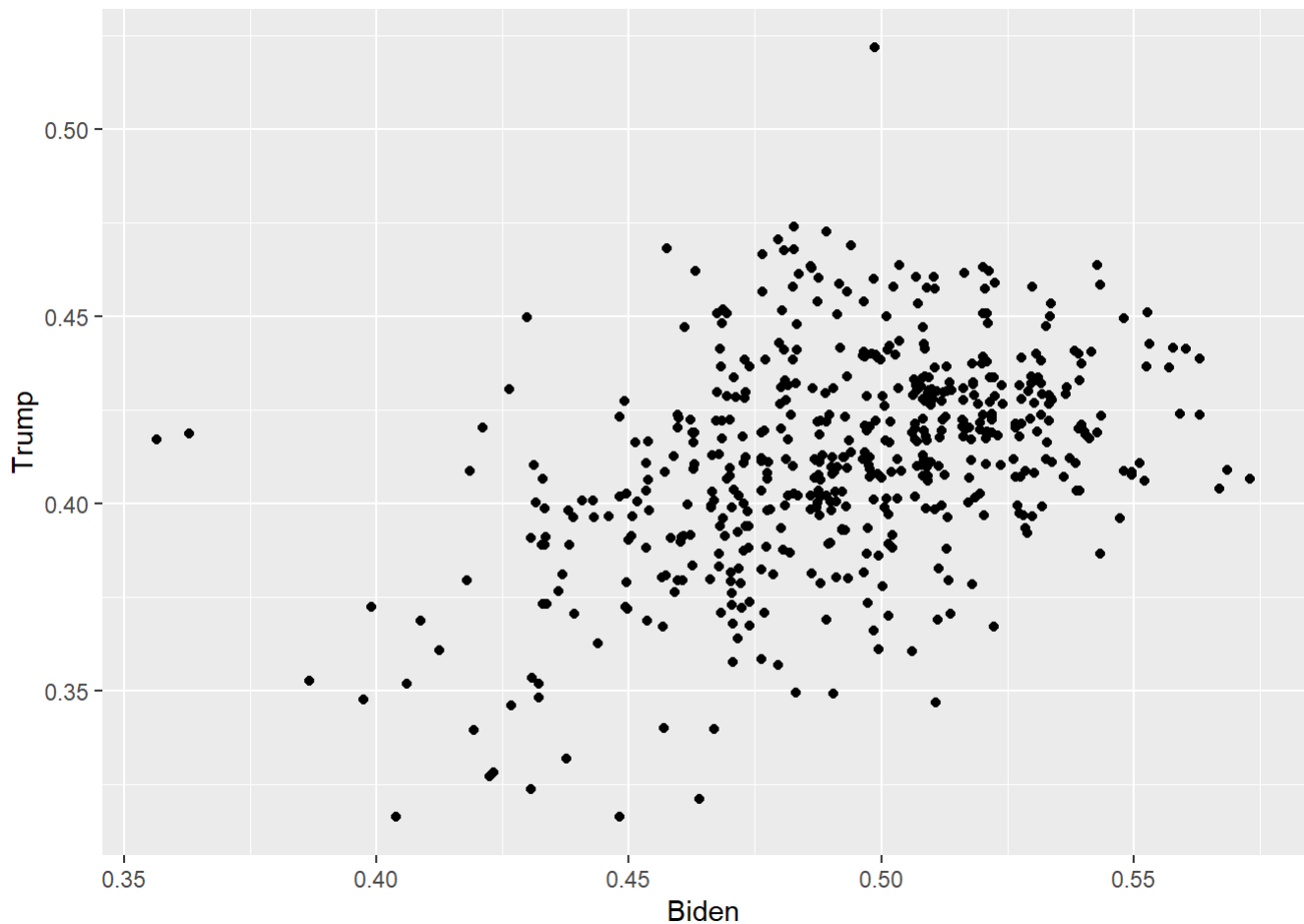
```
poll %>%  
  ggplot(aes(x = Biden, y = Trump)) +  
  geom_point()
```



```
# Illustrating multiple observations: alpha (opacity)
poll %>%
  ggplot(aes(x = Biden, y = Trump)) +
  geom_point(size = 3, alpha = .3)
```

```
# Illustrating multiple observations: geom_jitter()
poll %>%
  ggplot(aes(x = Biden, y = Trump)) +
  geom_jitter()
```



```
# Illustrating multiple observations: size
poll %>%
  count(Biden,Trump) %>%
  ggplot(aes(x = Biden,y = Trump,size = n)) +
  geom_point() +
  geom_smooth(method = 'lm',se = F)
```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: The following aesthetics were dropped during statistical transformation: siz
e
## i This can happen when ggplot fails to infer the correct grouping structure in
## the data.
## i Did you forget to specify a `group` aesthetic or to convert a numerical
## variable into a factor?
```

