QTM151exam

```
library(tidyverse)
```

```
## Loading tidyverse: ggplot2
## Loading tidyverse: tibble
## Loading tidyverse: tidyr
## Loading tidyverse: readr
## Loading tidyverse: purrr
## Loading tidyverse: dplyr
```

```
## Warning: package 'dplyr' was built under R version 3.4.2
```

```
## Conflicts with tidy packages ----------------------------------------------
```

```
## filter(): dplyr, stats
## lag():    dplyr, stats
```

```
library(tidyr)
library(dplyr)
library(ggplot2)
poll16<-read.csv("~/Downloads/poll16.csv")
```

Question 1:

Provide the appropriate graph to visualize the distribution of the total number of sample voters from each organization (the variable “Poll”), and find the top two organizations with the most total number of sample voters from the graph.

```
poll16_1<- poll16 %>%
  separate(Sample, into = c("sample", "voter"))
```

```
## Warning: Too few values at 1 locations: 33
```

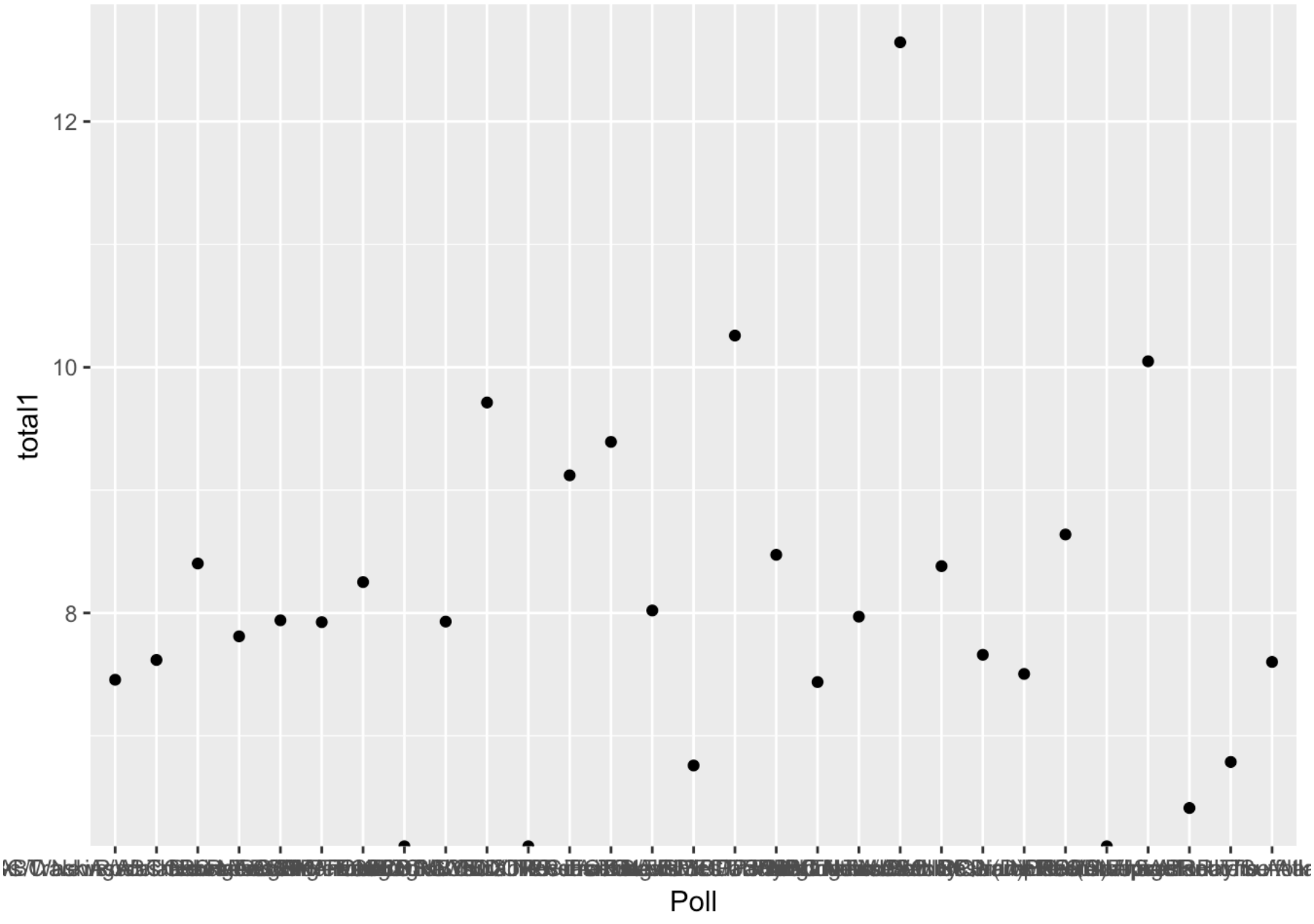
```
poll16_1$sample <- as.numeric(as.character(poll16_1$sample))
```

```
## Warning: NAs introduced by coercion
```

```
poll16_1 %>%
group_by(Poll) %>%
  summarise(total = sum(sample, na.rm=TRUE)) %>%
    arrange(desc(total)) %>%
   ungroup() %>%
      top_n(2, total)
```

```
## # A tibble: 2 x 2
##                            Poll  total
##                          <fctr>  <dbl>
## 1           NBC News/SMNBC News 310380
## 2 LA Times/USC TrackingLA Times  28509
```

```
poll16_1 %>%
group_by(Poll) %>%
  summarise(total1 = log(sum(sample, na.rm=TRUE))) %>%
ggplot(aes(x = Poll, y = total1)) +
  geom_point()
```



The top two organizations with the most total number of sample voters are NBC News/SMNBC News and LA Times/USC TrackingLA Times.

Question 2:

Provide the appropriate graph to visualize the difference between Clinton and Trump points by the top two organizations (you find in the question 1) at each day (use the end date).

```
poll16_2<- poll16_1 %>%
  separate(Date, into = c("startdate","enddate"),sep = "/")


poll16_2 %>%
  filter(Poll == "NBC News/SMNBC News" | Poll == "LA Times/USC TrackingLA Times") %>%
  mutate(difference = Clinton - Trump) %>%
ggplot(aes(x = enddate, y = difference))+
  geom_point(aes (color= Poll))
```