

LWV_HW8

Vinh Le

July 13, 2016

```
knitr::opts_chunk$set(echo=TRUE)
library(ggplot2)
```

The project spanned three semesters. In the spring of 2014, students estimated the sample size needed for the study and developed the experimental design and sampling methodology. In the fall of 2014, the data were gathered and cleaned, and in the spring of 2015, the data were analyzed. The sample consisted of 24,000 randomly selected individuals from a population of 531,735. The population was low-propensity voters, determined ahead of the 2014 election. A low propensity voter is someone who has only voted in 0 or 1 of the last three elections. The sample was randomly partitioned into three treatment groups.

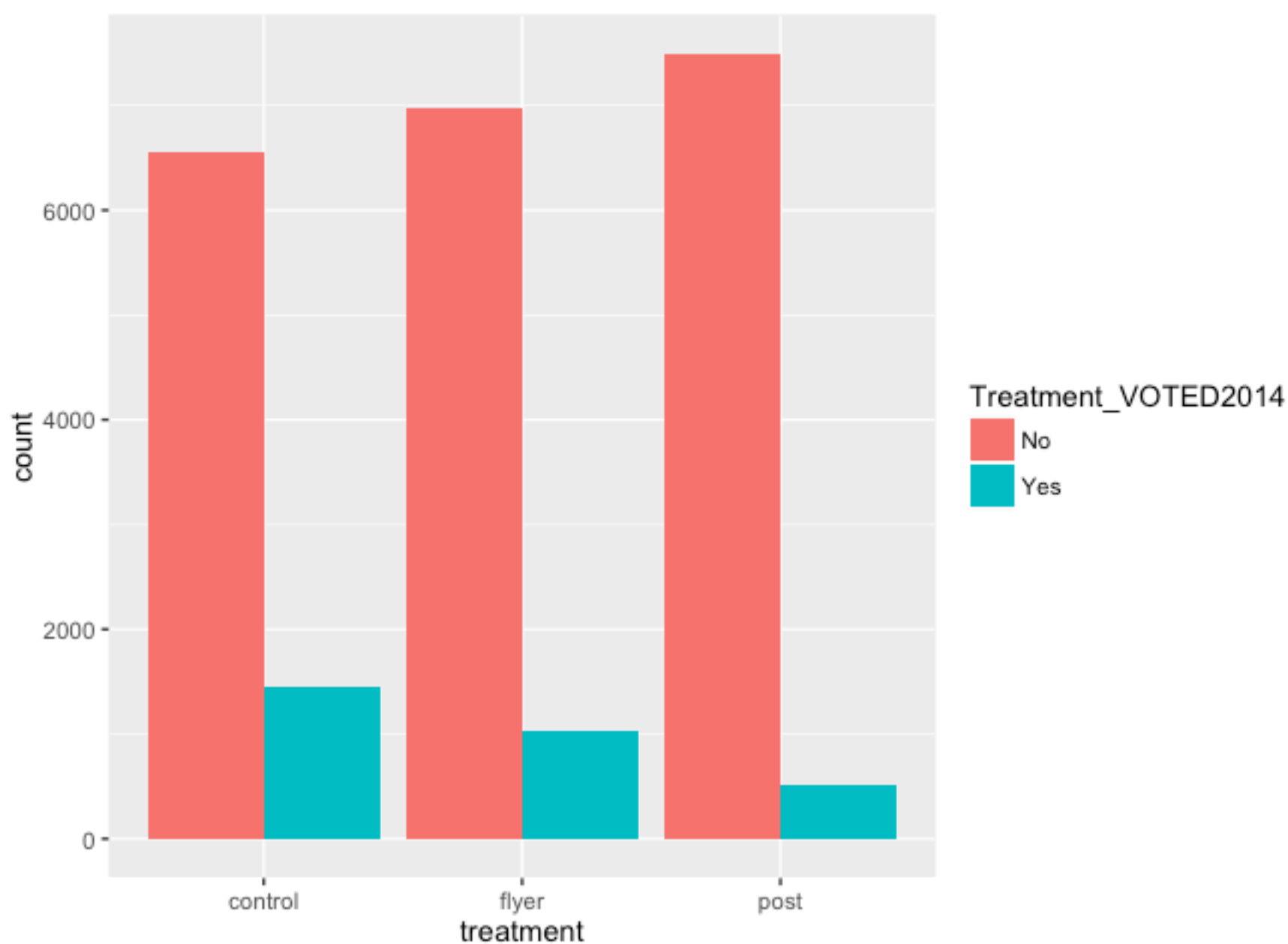
- 8000 received a postcard reminding them to vote
- 8000 received a flyer with a reminder and voting instructions
- 8000 did not receive any mailing. These are the control group.

After the 2014 election, information regarding voter participation was collected for all voters. When the data were analyzed, it was found that 10% of the low propensity voters who had received the flyer voted, approximately 12% of those who received the postcard voted, and approximately 23% of the control group voted.

```
data <- read.csv(file = "~/Desktop/SMU DATA SCIENCE/MSDS 6306 Intro to Data Science/Unit 8/Unit 8 HW/LWV_HW8/LWV_Data.csv", header = TRUE)
data$treatment[data$control == 1] <- "control"
data$treatment[data$post == 1] <- "post"
data$treatment[data$flyer == 1] <- "flyer"
data$Treatment_VOTED2014[data$VOTED2014 == 0] <- "No"
data$Treatment_VOTED2014[data$VOTED2014 == 1] <- "Yes"
```

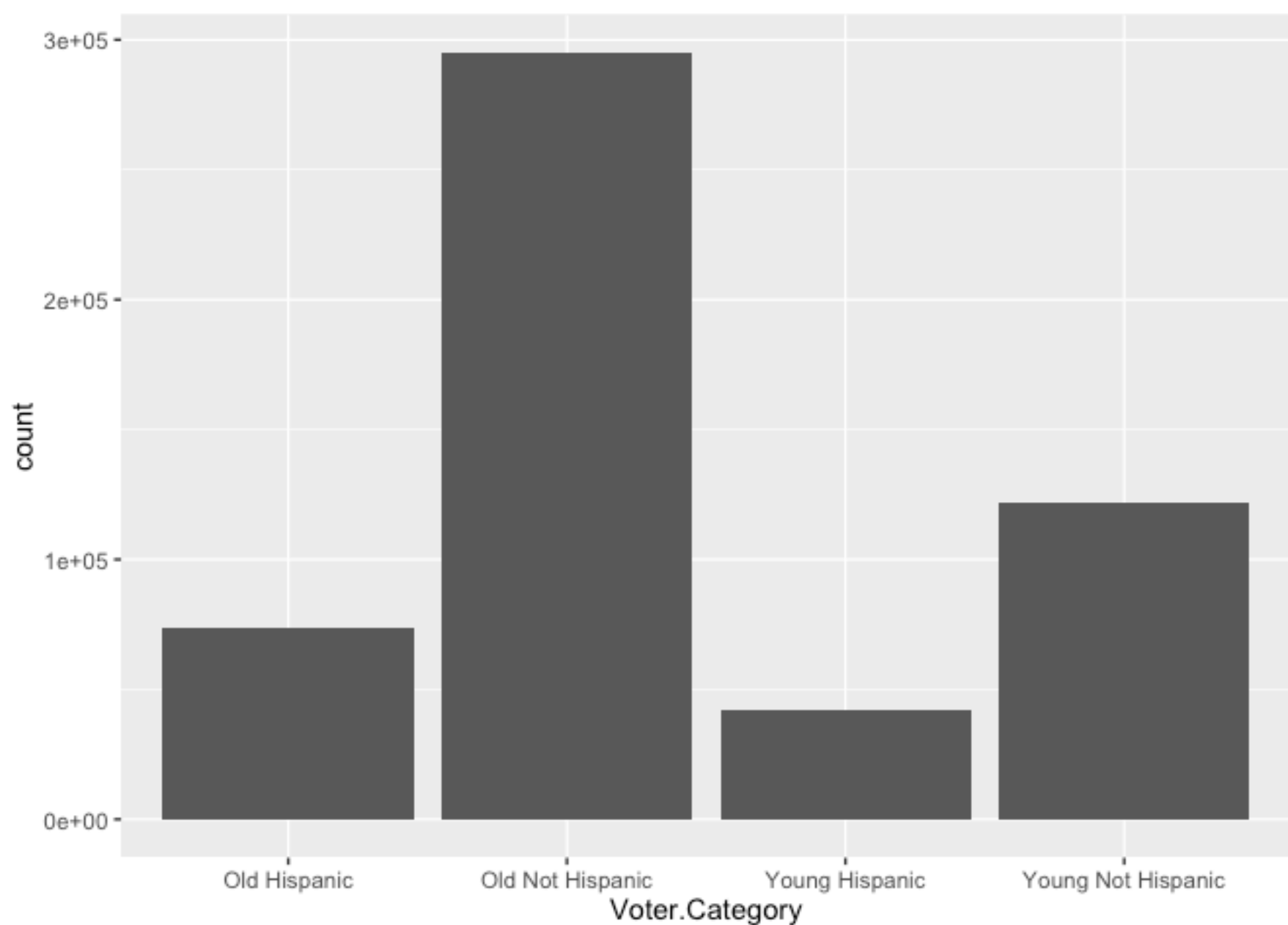
```
ggplot(subset(data), aes(x=treatment, fill=Treatment_VOTED2014))+geom_bar(aes(y= ..count..), stat="count", position = "dodge")
```

```
## Warning: Removed 507735 rows containing non-finite values (stat_count).
```



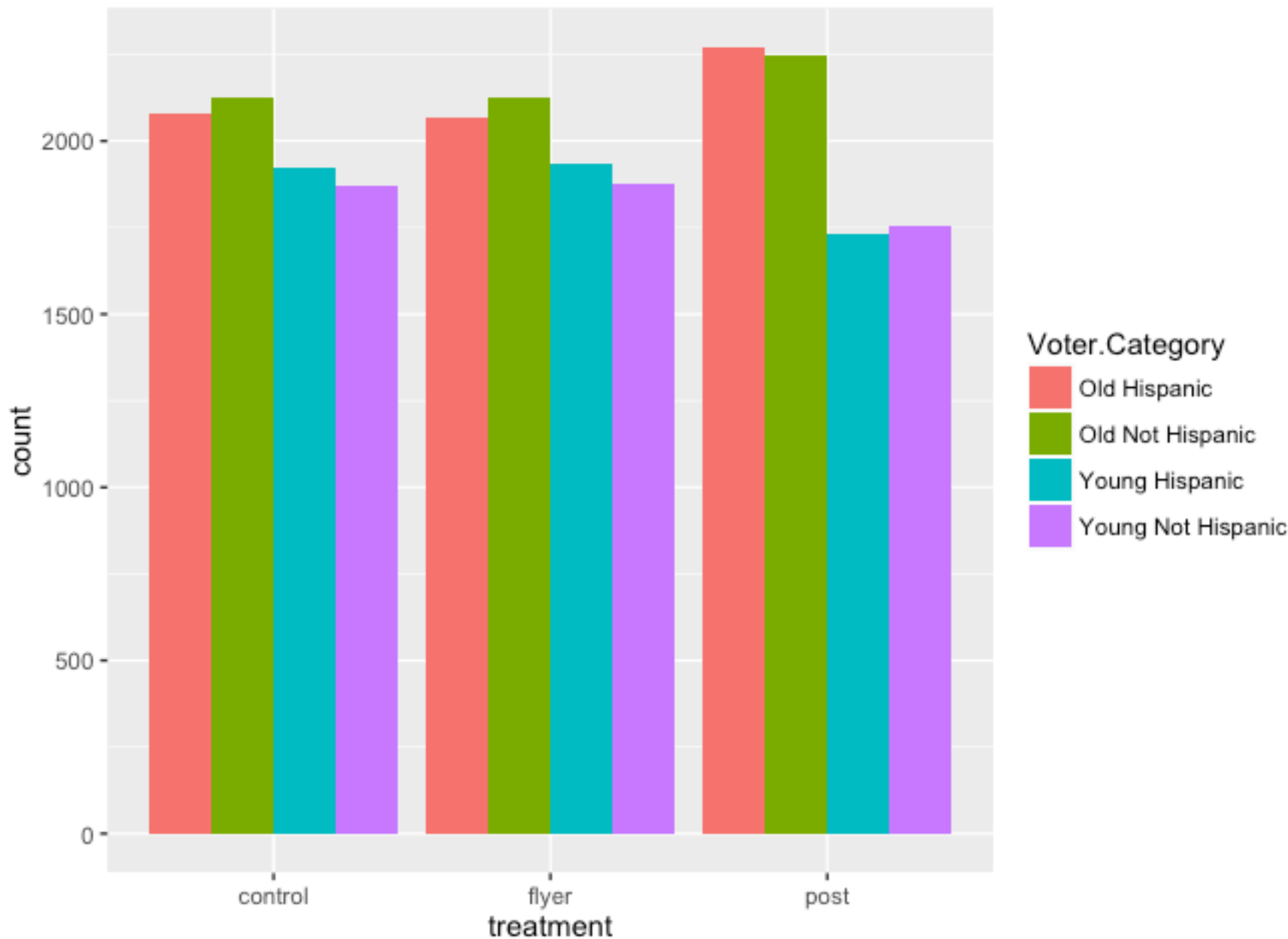
The graph shows that the control group had more votes than the other groups with reminder. We will investigate further to understand the reason behind it.

```
ggplot(subset(data), aes(x=Voter.Category))+geom_bar(aes(y= ..count..), stat="count")
```



```
ggplot(subset(data), aes(x=treatment, fill=Voter.Category))+geom_bar(aes(y= ..count..),  
stat="count", position="dodge")
```

```
## Warning: Removed 507735 rows containing non-finite values (stat_count).
```



Conclusion

We are concern why the control group had more votes. As you can see young people had less votes in all the treaments. There are more chances that old people will vote more so than young people regardless if there’s a reminder or not. Because the sample drawn from the population had an enormous amount of old people (both hispanic and non-hispanic) that in return would give more votes for the control group. Had the sample were evenly distributed between young and old and that the samples were more randomly selected then we would get a better reading.