HW #1

1.2, 1.6, 1.11, 1.12, 1.16, 1.18, 1.26, 1.27

Hayden Atchley

6 September 2022

1.2

The sex discrimination study is an observational study rather than a randomized study. Because of this, it is impossible to know what other factors may have contributed to the results other than the sex of the employees. Even if the researchers were to control for variables such as education, experience, age, and/or any other variables they could think of, there may still be confounding variables that aren't known and can't be accounted for. It is therefore difficult to prove sex discrimination from this study alone.

1.6

This study is an observational study, as it wasn't randomized which subjects used marijuana and which didn't. As such, it is difficult to conclude that marijuana use affects short-term memory because any number of other factors might be at play. The selection of the subjects was also not random, and so the results can't be generalized to a larger group.

Many other factors could potentially be confounding variables in this study. These could include education, genetics, factors from how they were raised, or potentially even indirect factors such as household income.

1.11

While at first glance it seems that the Vitamin C should have the same effect whether or not the participants know they're taking it, much research has shown that the "placebo effect", where the mere belief that a treatment will have a certain effect can cause that effect, is significant. The extra confidence of the participants who knew they weren't in the control group may have contributed to the results. Additionally, those participants may have (consciously or not) modified their behaviors in a way that lowered their chances of getting a cold. Therefore, it is no longer possible to conclude that the difference in cold rates is due solely to the Vitamin C.

1.12

The results, though they can't be generalized to a population due to the non-random sample, are still important because the treatment was randomized. The results therefore indicate that the treatment did have a significant effect on the subjects, and they indicate a need for future studies to determine if this effect can be generalized for a larger population.

1.16

I obtained the data for this problem from the `STAT 511 Class materials` BOX folder.

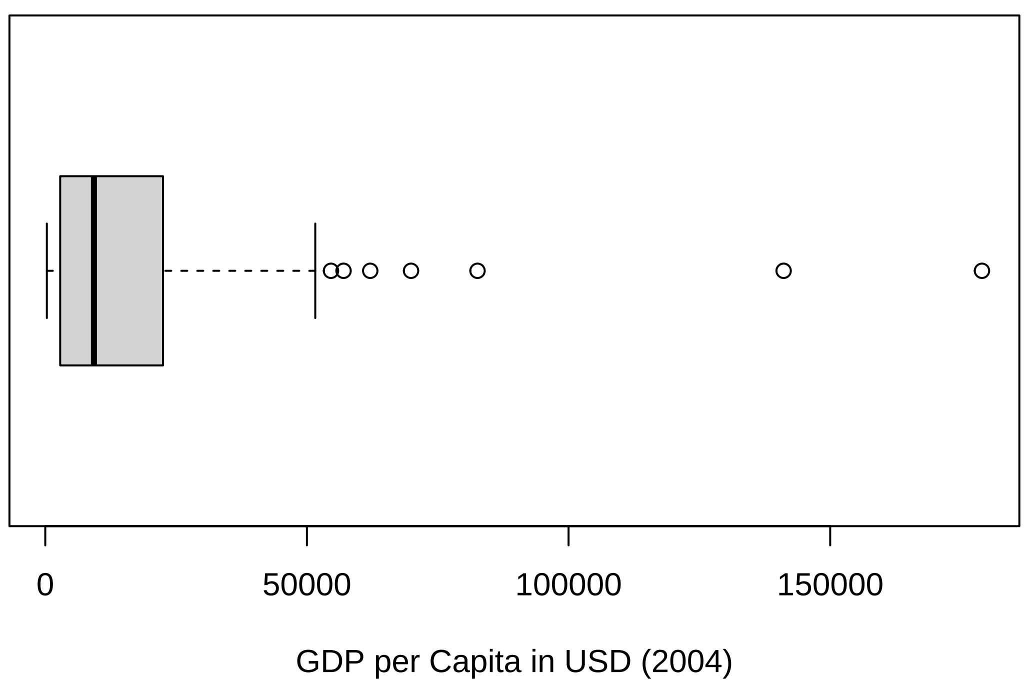
```
rank <- Sleuth3::ex0116

head(rank) %>%
    kable()
```

Rank	Country	PerCapitaGDP
1	Qatar	179000
2	Liechtenstein	141100
3	Luxembourg	82600
4	Bermuda	69900
5	Singapore	62100
6	Jersey	57000

A boxplot of the data is below.

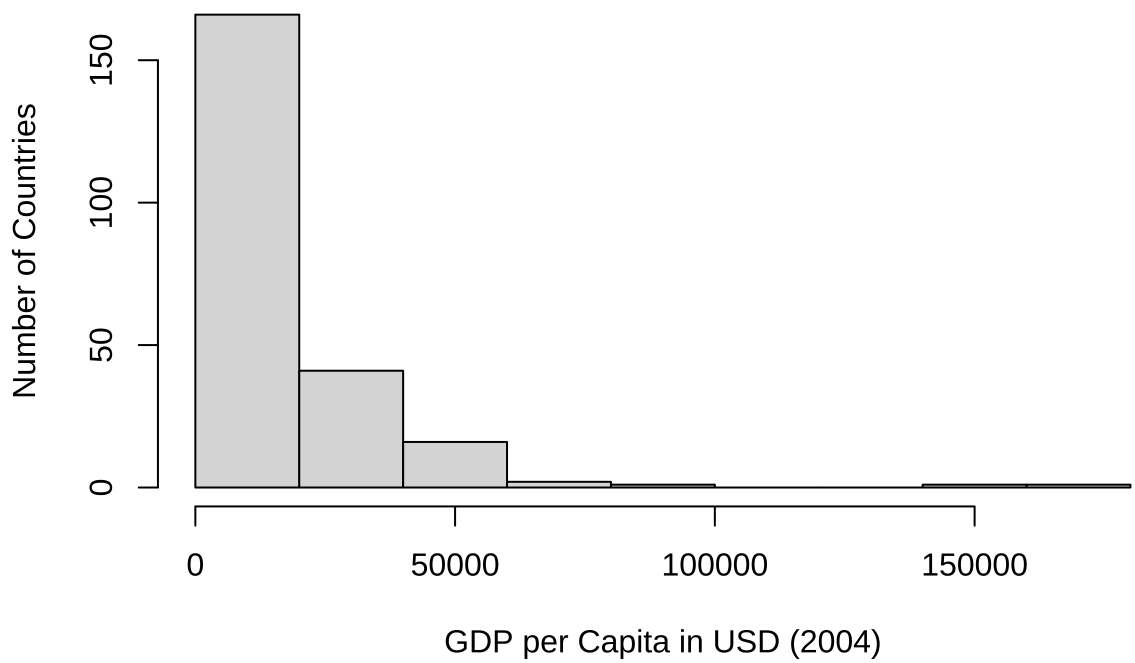
```
rank$PerCapitaGDP %>%
    boxplot(
        xlab = "GDP per Capita in USD (2004)",
        horizontal = TRUE)
```



This plot differs from Display 1.11 in several ways, mostly with presentation: Display 1.11 is vertical while this plot is horizontal, the Display names extreme outliers and labels various points of interest in the plot whereas this one does not, and the scale differs by a factor of 1000.

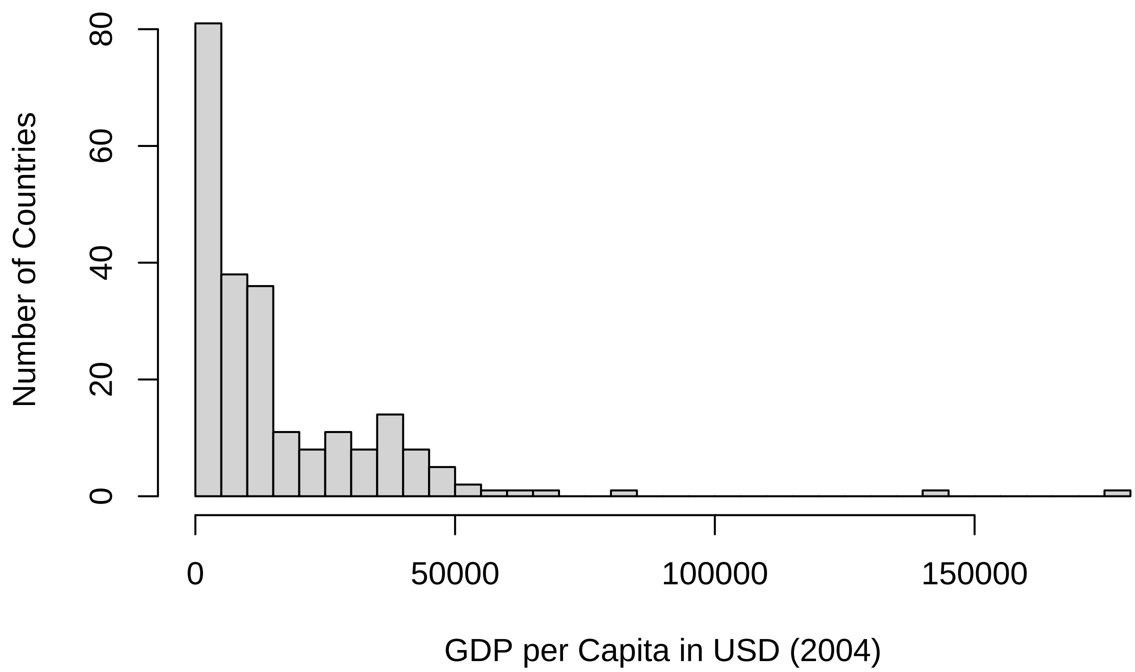
A histogram of the data is below. The bin width defaulted to $20000.

```
rank$PerCapitaGDP %>%
    hist(
        xlab = "GDP per Capita in USD (2004)",
        ylab = "Number of Countries",
        main = NULL)
```



Changing the bin width to $5000 gives the following:

```
rank$PerCapitaGDP %>%
    hist(
        xlab = "GDP per Capita in USD (2004)",
        ylab = "Number of Countries",
        breaks = seq(
            RoundTo(min(.), 5000, "floor"),
            RoundTo(max(.), 5000, "ceiling"),
            5000),
        main = NULL)
```



1.18

```
Intrinsic <- Sleuth3::case0101 %>%
    filter(Treatment == "Intrinsic") %>%
    {.$Score}
Extrinsic <- Sleuth3::case0101 %>%
    filter(Treatment == "Extrinsic") %>%
    {.$Score}

obsdiff <- mean(Intrinsic) - mean(Extrinsic)

tibble(
    "Group" = c("Intrinsic", "Extrinsic"),
    "Sample Size" = c(length(Intrinsic), length(Extrinsic)),
    "Mean" = c(mean(Intrinsic), mean(Extrinsic)),
    "Standard Deviation" = c(sd(Intrinsic), sd(Extrinsic))) %>%
    kable(digits = 2)
```

Group	Sample Size	Mean	Standard Deviation
Intrinsic	24	19.88	4.44
Extrinsic	23	15.74	5.25

Running a two-tailed randomization test:

```
randtest <- twoSamplePermutationTestLocation(
    Intrinsic,
    Extrinsic,
    n.permutations = 1000)
```

The test gives 1000 observations of the difference in mean for the randomized dataset. The observed difference in mean was 4.14, and 2 of these randomized differences were greater than the observed difference:

```
randtest$stat.dist %>%
    {.[. > (mean(Intrinsic) - mean(Extrinsic))]} %>%
    round(2)
```

```
## [1] 4.43 4.21
```

This gives a 2/1000 probability that the observed difference in means is by chance, or 0.002.

1.26

First I loaded the data. Because there were only 3 individuals not associated with the Republican or Democratic party, I removed those data points.

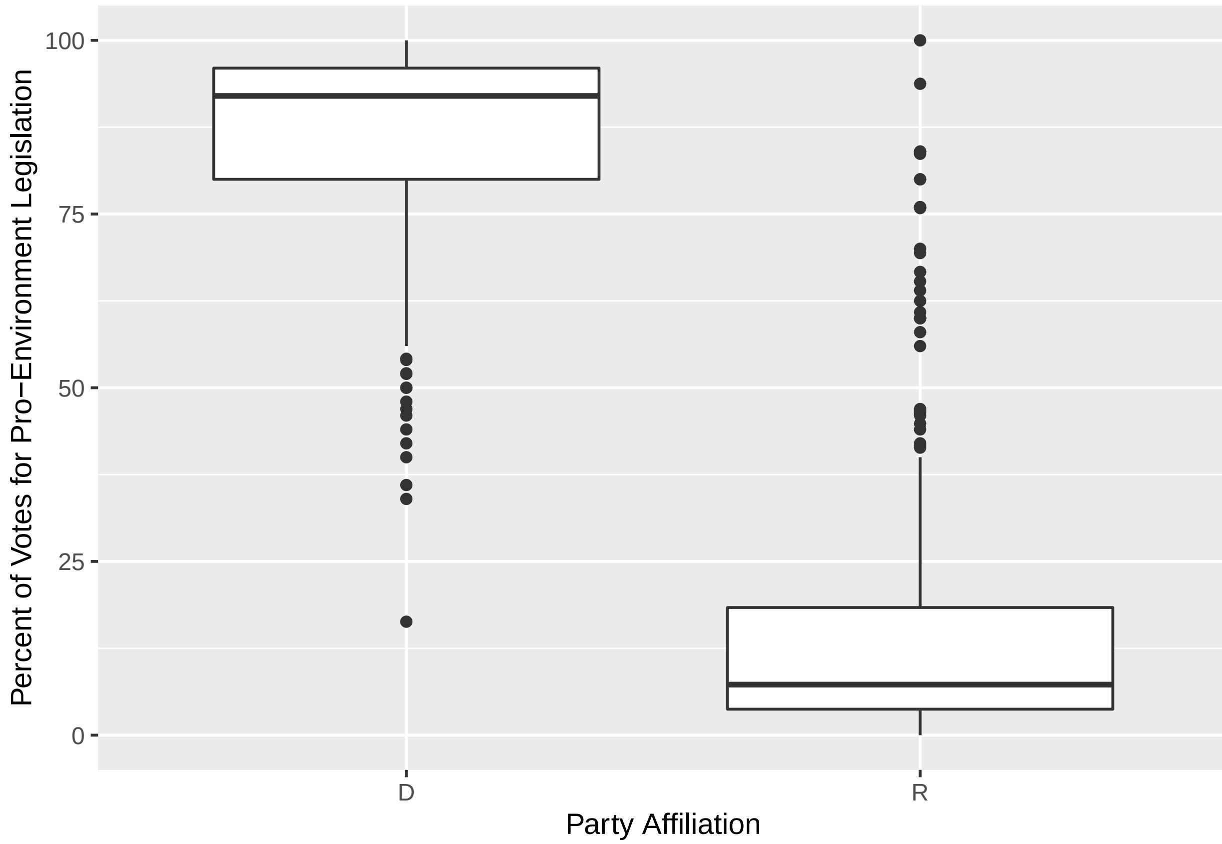
```
votes <- Sleuth3::ex0126 %>%
    filter(Party %in% c("R", "D"))

head(votes) %>%
            kable()
```

State	Representative	Party	Pro05	Anti05	Pro06	Anti06	Pro07	Anti07	PctPro
Alabama	Bonner	R	2	16	3	9	2	18	14.00
Alabama	Everett	R	0	18	1	11	2	18	6.00
Alabama	Rogers	R	1	17	2	10	3	17	12.00
Alabama	Aderholt	R	0	18	0	12	2	18	4.00
Alabama	Cramer	D	5	13	4	7	14	6	46.94
Alabama	Bachus	R	0	18	1	11	1	19	4.00

To get an idea of the data, I made a boxplot comparing the two parties:

```
votes %>%
    ggplot(aes(x = Party, y = PctPro)) +
    geom_boxplot() +
    labs(
        x = "Party Affiliation",
        y = "Percent of Votes for Pro-Environment Legislation")
```

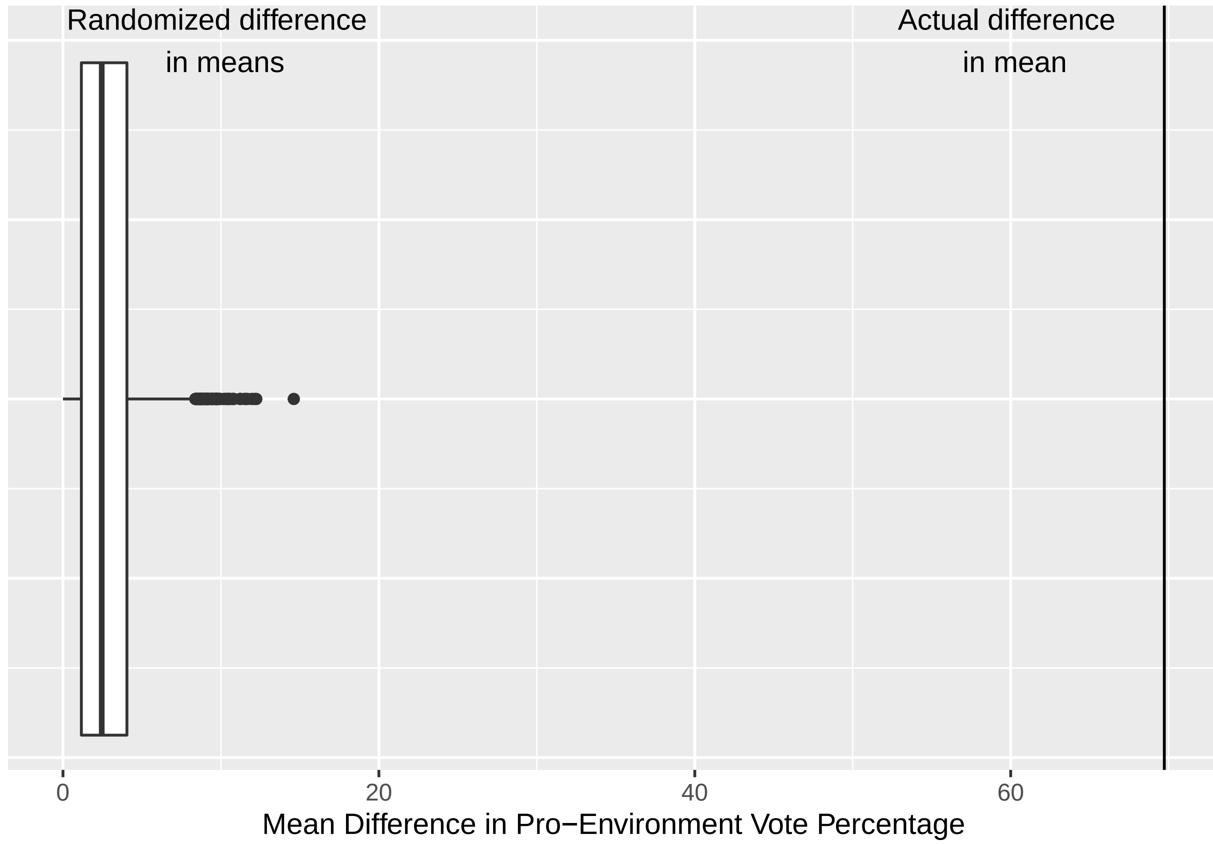


This plot is somewhat striking in itself, but I also did a randomization test:

```
Rep <- votes %>%
    filter(Party == "R") %>%
    {.$PctPro}
Dem <- votes %>%
    filter(Party == "D") %>%
    {.$PctPro}

randtest <- twoSamplePermutationTestLocation(Rep, Dem)
```

The observed difference in mean is 69.73, but none of the randomized differences even come close to that number:

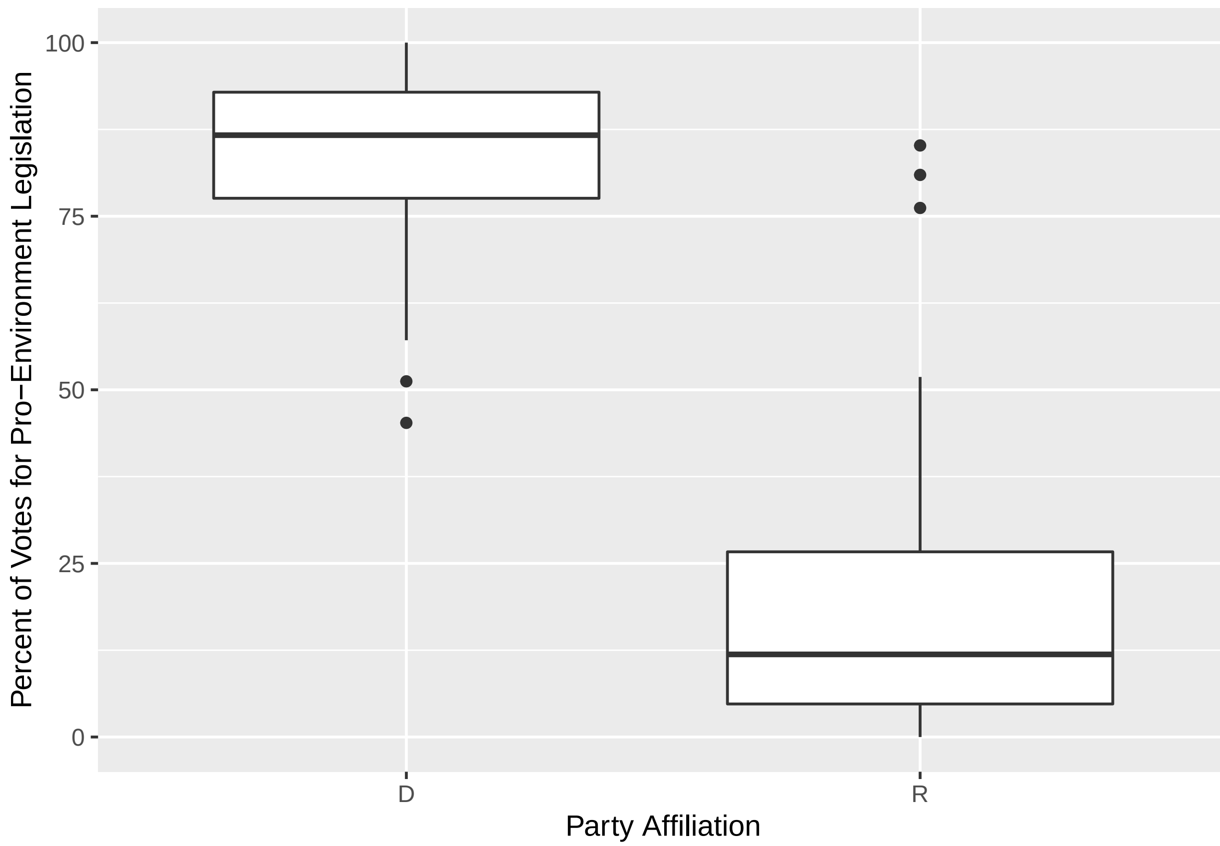


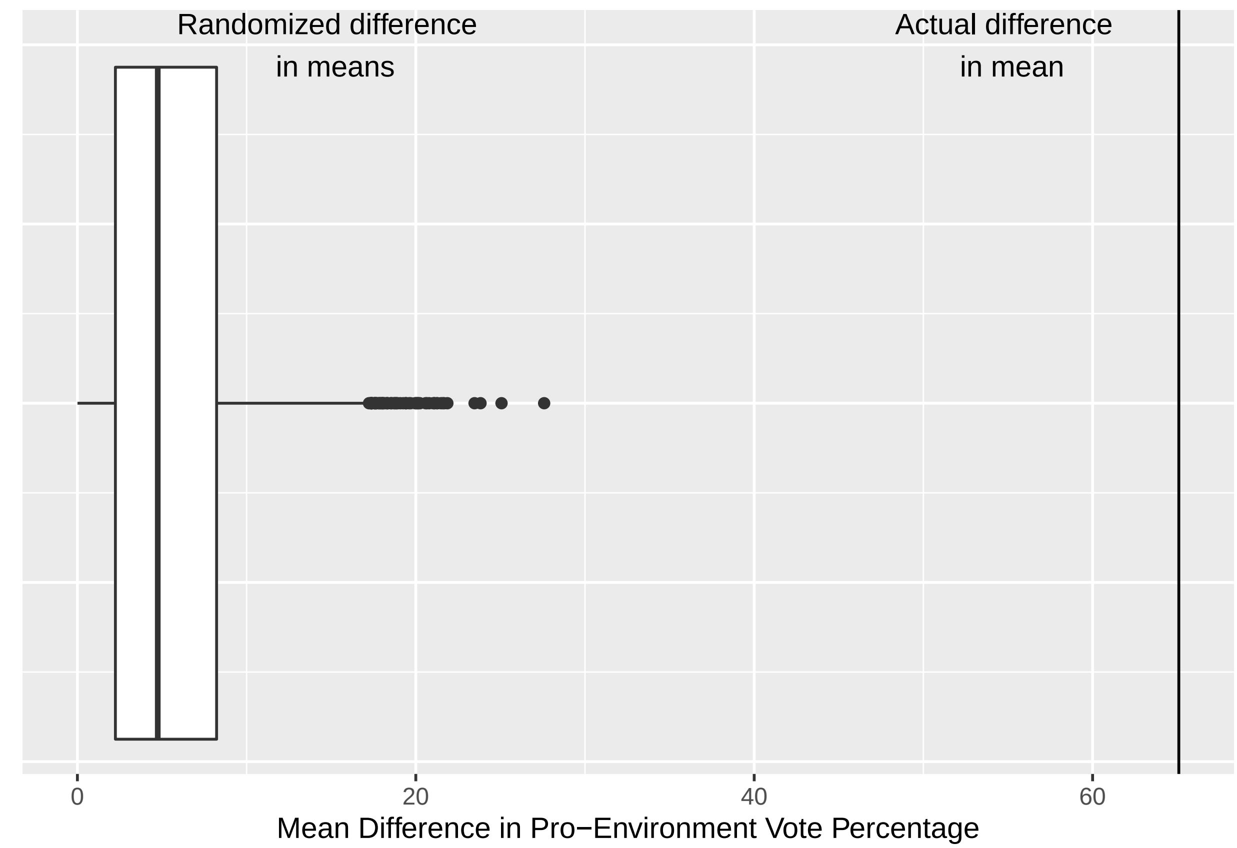
It is clear that this difference is not by chance.

1.27

This problem is nearly identical to the previous one, but with the Senate rather than the House.

State	Senator	Party	Pro2005	Anti2005	Pro2006	Anti2006	Pro2007	Anti2007	PctPro
Alabama	Session	R	1	18	0	7	2	13	7.32
Alabama	Shelby	R	1	19	0	7	1	14	4.76
Alaska	Murkowski	R	2	17	1	6	6	9	21.95
Alaska	Stevens	R	1	19	1	6	4	11	14.29
Arizona	Kyle	R	1	19	2	5	2	13	11.90
Arizona	McCain	R	9	11	2	4	0	15	26.83





This difference is also clearly not by chance.