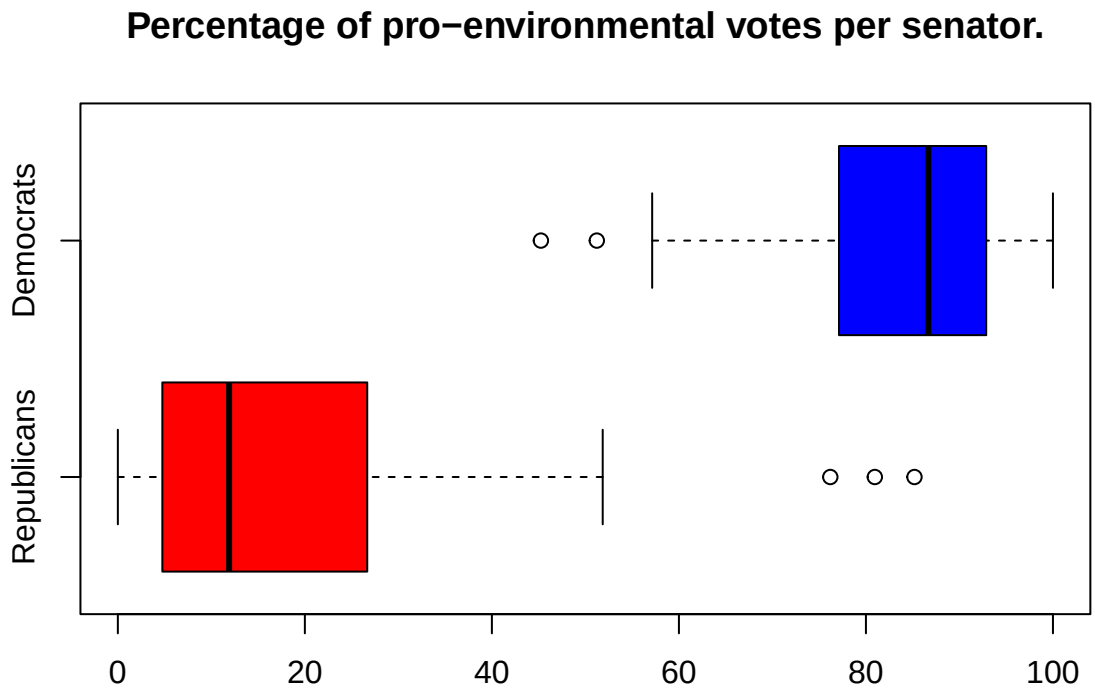
Week 02 solutions

2023-02-08

Part one: Tuesday

1. Environmental Voting of Democrats and Republicans in the U.S. Senate. In this exercise we will consider a data set according to the League of Conservation Voters of pro- and anti-environmental votes cast by U.S. senators during 2005 - 2007. Use visual representation of the data to investigate party differences in the percentage of pro-environment votes.
   Solution:

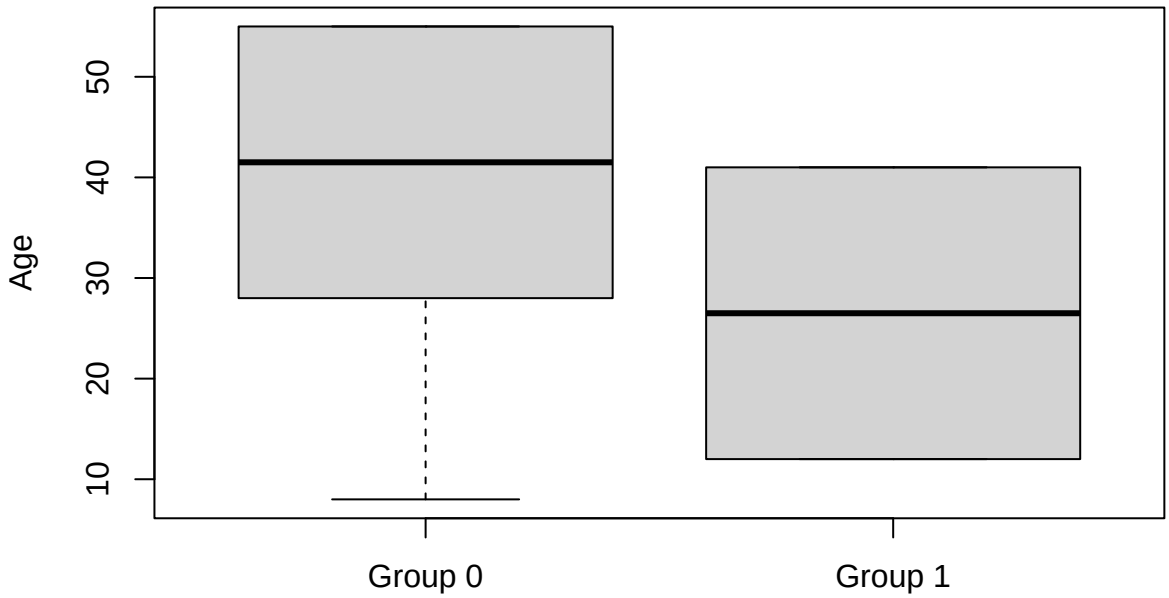
```
vote_data <- read.csv('data/Environmental votes.csv')
republicans <- dplyr::filter(vote_data, Party == "R")
democrats <- dplyr::filter(vote_data, Party == "D")
boxplot(republicans$PctPro, democrats$PctPro,
        main="Percentage of pro-environmental votes per senator.",
        names=c("Republicans", "Democrats"),
        col=c("Red", "Blue"),
        horizontal=TRUE)
```



2. Fish Oil and Blood Pressure. Consider the study from previous week about the correlation between a diet containing fish oil and blood pressure. In the study, the researchers used 7 red and 7 black playing cards to randomly the participants to the treatment groups. Does this method constitute a random sampling? Why might the results of this study be important?
   Solution:
   This depends on how the participants were randomized, whether the cards were shuffled etc. If we assume the cards were fairly shuffled, and the participants each randomly chose a card, it is a random sample.
   The results are important as they suggest consuming fish oil reduces blood pressure, which could be a potential treatment option with more research.

3. Write down the names and ages of 10 people. Using coin flips, divide them into two groups, as if for a randomized experiment. Did one group tend to get many of the older subjects? Was there anyway to predict which group would have a higher age in advance of the coin flips?
   Solution:

```
people <- data.frame(
  name=c("Gideon", "Maddison", "Alisha", "Brodie", "Cian", "Osman", "Ciara", "Alesha", "Jesse", "Kurtis
  age=floor(runif(10, 1, 60)),
  group=rbinom(10, 1, 0.5)
)
boxplot(
  dplyr::filter(people, group == 0)$age, dplyr::filter(people, group == 1)$age,
  names=c("Group 0", "Group 1"), ylab="Age"
)
```



It is not possible to predict which group will have a higher age, as the distribution is random.

4. Zinc concentration. An experiment was conducted where a group of rats were split into two groups. One group, group A, received a dietary supplement, while the other group, group B, did not. We are interested in the possible side effect of the treatment. In the data set, the concentration of zinc (in mg/ml) in each rat is listed, as well as the rats' treatment group. Make box-plots for each of the two groups of rats. Preferably in the same plot, and if not then at least with the scale of the axis.
Solution:

```
rat_data <- read.csv('data/Zinc concentration.csv')
boxplot(
  dplyr::filter(rat_data, Group == "A")$Zinc, dplyr::filter(rat_data, Group == "B")$Zinc,
  names=paste("Group", c("A", "B")), xlab="Zin concentration (mg/ml)",
  horizontal=TRUE,
  col=c("darkorchid4", "darkorchid1")
)
```