

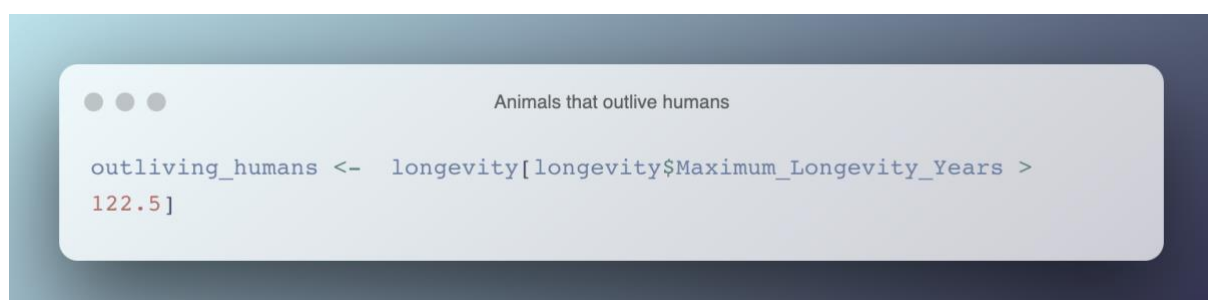
Objects, Class, and Data Structures

- *All materials for the exercises below are available in the homework folder.*
- *Please submit an R script file containing the code and results.*
- *You can #comment out any sentences written answers.*

*

Exercise 1: Animal longevity dataset

1. The animal longevity dataset contains data on the maximum (known) lifespan of various animals.
 - 1.1. How many variables and how many observations are there?
 - 1.2. Pick three animals from the dataset. Create three objects (named after the animal) storing the value of their longevity.
 - 1.3. Create three logical tests comparing the longevity between the objects you created in question 1.2.
 - 1.4. Create two vectors:
 - 1.4.1. The names of all the animals you chose in question 1.2.
 - 1.4.2. Their respective lifespans.
 - 1.5. Calculate the mean and median of their lifespans.
 - 1.6. Retrieve the 2nd value of the vector that contains your animal names.
 - 1.7. Subset which animals in your lifespans vector have a maximum longevity of above 27.
 - 1.8. Create a new dataset containing the maximum longevity of animals and their common names ranked from shortest to longest.
 - 1.9. What is the mean and median of maximum longevity of animals? Briefly comment these results.
 - 1.10. Suppose you want to know which animals can outlive humans. According to the dataset (longevity), the maximum longevity for humans (Maximum_Longevity_Years) is 122.5. Will the following code work?



```
Animals that outlive humans

outliving_humans <- longevity[longevity$Maximum_Longevity_Years >
122.5]
```

Exercise 2: Self-reported Turnout

2. Surveys are frequently used to measure political behavior such as voter turnout, but some researchers are concerned about the accuracy of self-reports. They worry about possible social desirability bias where, in postelection surveys, respondents who did not vote in an election lie about not having voted because they may feel that they should have voted. In this exercise we will try to identify such a bias in ANES.

ANES is a US-wide survey that has been conducted for every election since 1948. ANES is based on face-to-face interviews with a nationally representative sample of adults. The table below displays the names and descriptions of variables in the turnout.csv data file.

Variable	Description
Year	Election year
ANES	Self-estimated turnout rate
VEP	Voting eligible population (in thousands)
VAP	Voting age population (in thousands)
Total	Total ballots cast for highest office (in thousands)
Felons	Total ineligible felons (in thousands)
Noncitizens	Total noncitizens (in thousands)
Overseas	Total eligible overseas voters (in thousands)
Osvoters	Total ballots count overseas voters (in thousands)

- 2.1. Load the data into R and check the dimensions of the data. Also, obtain a summary of the data. How many observations are there? What is the range of years covered in this data set?
- 2.2. Calculate the real turnout rate based on the voting age population (VAP) by year. This is, how many voters actually voted ($VAP/TOTAL$) in each year. Note that for this data set, we must add the total number of eligible overseas voters since the VAP variable does not include these individuals in the count. Next, calculate the real turnout rate using the voting eligible population (VEP). This is $VEP/TOTAL$. What difference do you observe?
- 2.3. Compute the differences between the self-estimated turnout rate (ANES) and the real turnout rate you just calculated in 2.2. How big is the difference on average? Conduct the same comparison for the VEP and ANES estimates of voter turnout. Briefly comment on the results.
- 2.4. Compare the VEP turnout rate with the ANES turnout rate separately for presidential elections and midterm elections. Note that the data set excludes the year 2006. Does the bias of the ANES estimates vary across election types?
- 2.5. Divide the data into half by election years such that you subset the data into two periods: older and newer elections. Calculate the difference between the VEP turnout rate and the ANES turnout rate separately for each year within each period. Has the bias of ANES increased over time?