Week 3 Tutorial Exercises

Andrew Perfors

We’ve seen from Bunny and Gladly’s surveys so far a kind of weird and maybe interesting thing: when asked to rate how much they’d want to eat various items right now, all of the ratings for food items were very high, with averages and means well above 5 on a 10-point scale and very few ratings below 5 at all. We know it’s not because people just don’t understand the questions, because the ratings for questions about non-food items like mud are exactly where you’d expect: around 1 or 2.

So… what’s going on?

“I think it’s just that food is very yummy and people like to eat all of the time,” says Gladly.

“That doesn’t make sense though,” says Shadow. “Surely they would rate things pretty low after they’ve just eaten dinner?”

“That’s a good idea,” says Doggie. “Maybe you asked all of the questions right before dinner or something, Gladly.”

“No,” says Gladly. I usually asked them right after breakfast in fact, because that’s when everybody is easy to find.”

“I’m worried,” says Bunny.

“Yeah,” says Flopsy. “What if we don’t have enough food? What if everybody is starving?”

“Oh calm down, Bunny,” Doggie says, irritated. (He is often irritated by Bunny’s anxiety, and not very good at hiding it). “Don’t you think we’d realise if we were starving?”

“Well… I *am* hungry all of the time,” says Gladly.

“Of course you are, you’re a bear!”

“Please stop bickering,” Bunny says miserably. “It makes me sad.”

Everyone is quiet for a moment, and then Shadow says, “Hey, I have an idea! We can use data to figure this out.”

“Where can we get more data?” asks Flopsy. “It’s not like it’s just lying on the ground.”

“I just remembered that LFB and Foxy have been doing a yearly survey on just this question,” Shadow says. “Quackers was helping them.”

“Oh yeah that’s right!” says Gladly. “That’s where I got the idea for the questions in the first place.”

“Was Quackers really helping them?” asks Bunny skeptically. “All he does is make jokes. And Foxy is so quiet, I can’t believe she could stand up to him.”

Flopsy says, “You might be surprised, I think. Foxy is just shy because she’s not a native speaker of Bunnish. She’s pretty strong though.”

“*Anyway*”, interrupts Shadow impatiently, “Let’s ask them for their data. We can see if anything has changed over time. Also I think they asked about more foods, so we can see if it’s about that.”

The friends all agree that that was a great idea, and get that dataset. In this tutorial activity you’ll use your skills to analyse it and see if you can figure out what is going on. Are the bunnies and friends slowly starving to death, or just bad at using the scales on this survey? Has anything changed over time? These are the questions for today.

# Q1

First, load the libraries and data you’ll need. I’ve filled in some of the libraries but you may find as you go through that you need to add them here. The dataset is called shadowsurvey.csv and it can be found in this folder, assuming you didn’t move anything when you downloaded the ZIP file from Canvas and are working out of the unzipped folder.

# We'll begin by loading up the libraries and data we need, as always.  
knitr::opts\_chunk$set(echo = TRUE)  
  
# loading the libraries  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(here)

## here() starts at /Users/henrikformoe/Desktop/RMHI\_tutorials/tuteweek3

# put in commands to load the dataset into a tibble called d

## Description of the dataset

This dataset is very much like Gladly’s which we’ve spent so much time in lecture on, except that LFB and Foxy have been collecting this data for five years and also asked about more foods than Gladly did. The columns in the dataset d are:

*name*: name of the person

*gender*: their reported gender

*species*: their species

*year*: the year the question was asked

*carrot,bean,cake,meat,mud,dirt*: questions about how much the person in question wanted to eat each of these items when asked, rated on a scale from 1 to 10

# Q2

Let’s first get a sense of the dataset. Use the summary command to do so.

# Q3

It’s really annoying to have all six of these survey questions when really there are only two main distinctions we want to make: between the foods we expect people to rate highly (carrot, cake, bean, and meat) and the things that aren’t food at all and we expect them to rate low (mud, dirt). To that end, make two new variables in d called goodFood and badFood. goodFood should be the mean of the responses to the four foods and badFood should be the mean of the responses to mud and dirt. Note: you can either do this using the techniques discussed in Week 2, or if you’ve already seen the videos for Day 2 in Week 3, using the mutate operator.

# Q4

Make two histograms, one for each of the two new variables you’ve made (goodFood and badFood). Make sure you have a clear title for each, and if you want to, set a different colour for each too. What do these histograms show you?

# Q5

Calculate the mean, median, mode, and 10th and 90th quantile of goodFood. Don’t use summary(), use the functions for each. (Note: you may need to load another library). Do these numbers make sense in light of the histograms?

# Q6

One of the different things about this dataset (compared to Gladly’s) is that there is data from multiple years. So let’s investigate it! Make a frequency table showing the number of data points for each year. How is it changing over time? Why do you suppose it is?

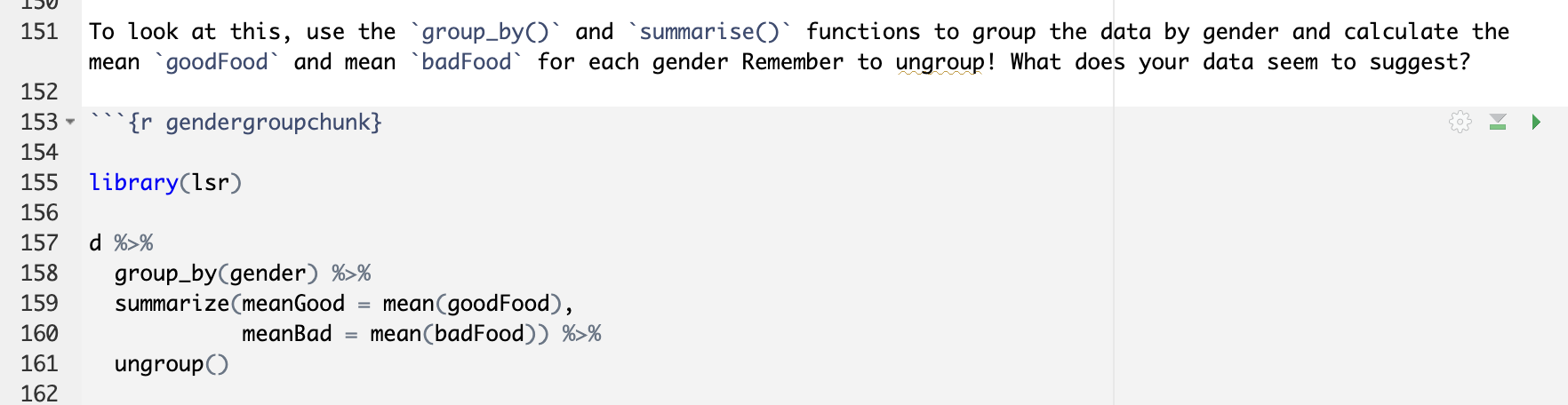
# Q7

Let’s get a breakdown of what species are represented each year. Make a table of species by year. The columns should be the years and the rows the species.

# Q8

Let’s first see if there are gender differences in how people anv1wer the questions.

To look at this, use the group\_by() and summarise() functions to group the data by gender and calculate the mean goodFood and mean badFood for each gender Remember to ungroup! What does your data seem to suggest?



# Q9

Another interesting thing to look at would be how the answers to the questions change over time. To look at this, use the group\_by() and summarise() functions to group the data by year and calculate the mean and sd for goodFood, the mean and sd for badFood, and the n for each year. Remember to ungroup! What does your data seem to suggest?

library(lsr)

d %>%

group\_by(gender) %>%

summarize(meanGood = mean(goodFood),

meanBad = mean(badFood)) %>%

ungroup()

library(lsr)

d %>%

group\_by(year) %>%

summarize(meanGood = mean(goodFood),

meanBad = mean(badFood)) %>%

summarize(sdGood = sd(goodFood),

sdBad = sd(badFood) %>%

ungroup()

# Q10

Knit your document to Word. (You’ll want to practice this with every tutorial document because the assignment will involve knitting to Word and you don’t want to be doing it for the first time on the assignments.)