DATA VISUALIZATION

There are four rules that apply to all projects so far:

a) Follow instructions *precisely*. If I do not tell you what to write on a particular line, leave it blank.

b) Do not use any functions or approaches to problems that we have not yet learned in this course.

c) All code must be *scalable by sample size* unless specifically noted otherwise. This means your code should work equally well on a dataset with N=10 as N=1000.

d) Any code using *magrittr* should contain a max of one verb per line.

This week, you'll be visualizing datasets from a past project. All plots should be created using *ggplot2* unless otherwise specified. When linking portions of a ggplot(), use separate lines for each new + phrase. Download week3.csv from Project 3 into your data folder.

Warning. If you find that your graphs aren't what you expect from commands you think are correct, you probably messed up something in step 5.

Part 1 – Set up a new R Studio Project with one R script called week7.R

Part 2 – Data Import and Cleaning

1. **Lines 1-3:** Write a comment that says: **R Studio API Code**, and set the wd as usual.
2. **Line 5:** Write a comment that says: **Libraries**
3. **Lines 6-9:** Import any libraries you need for the remainder of this project here.
4. **Line 10:** Write a comment that says: **Data Import and Cleaning**
5. **Line 11-17**: Using a single pipe, (11) import the week3.csv dataset as a tbl called *week7_tbl*, (12-15) convert any data types that were not imported correctly into their correct types using tidy functions where possible and relabel factors according to Table 1, (16) remove all cases that failed the manipulation check (i.e., responded with anything other than a 1 to q6), and (17) drop q6 from the final tbl.

Part 3 – Visualization

6. **Line 19:** Write a comment that says: **Visualization**
7. **Line 20:** Create a single figure that displays a scatter plot and correlation matrix of q1-q5 and q7-q10 as well as density plots (i.e., the line graph version of a histogram). Do not use **ggplot**().
8. **Line 21-24:** Check for differences in Q1 over time by creating a plot of timeStart on the x-axis and q1 on the y-axis that looks like Fig 1.
9. **Line 25-26:** Check for gender differences in the Q1/Q2 relationship by creating a figure that looks like Fig 2. Note that this is a jitter plot, so your dot positions will probably not look the same.
10. **Line 27-31:** Check for gender differences in the Q1/Q2 relationship again, but this time by creating two side-by-side plots as shown in Fig 3 (also a jitter plot).
11. **Line 32-36:** Using a single pipe, create a plot displaying the experimental time elapsed in seconds between timeStart and timeEnd split by gender, which looks like Fig 4.
12. **Line 37+:** Recreate Fig 5. Jitter may be slightly different but should be similar. The background of the legend is 12.5% grey. *Hint*: Tackle this figure one plus at a time, checking your work as you go.

Table 1

Variable	Level	Label
Condition	A	Block A
Condition	B	Block B
Condition	C	Control
Gender	M	Male
Gender	F	Female

Note for figures. Figures may look wider or narrower due to R Studio settings.
Don't worry about this. Just be sure the features are the same. Also note that the
ggplot2 default theme has changed between when your Data Camp lessons were
created and now. If you do not update *ggplot2*, it will be harder to create these figures
(and will require more lines than you have available). Be sure to update first.
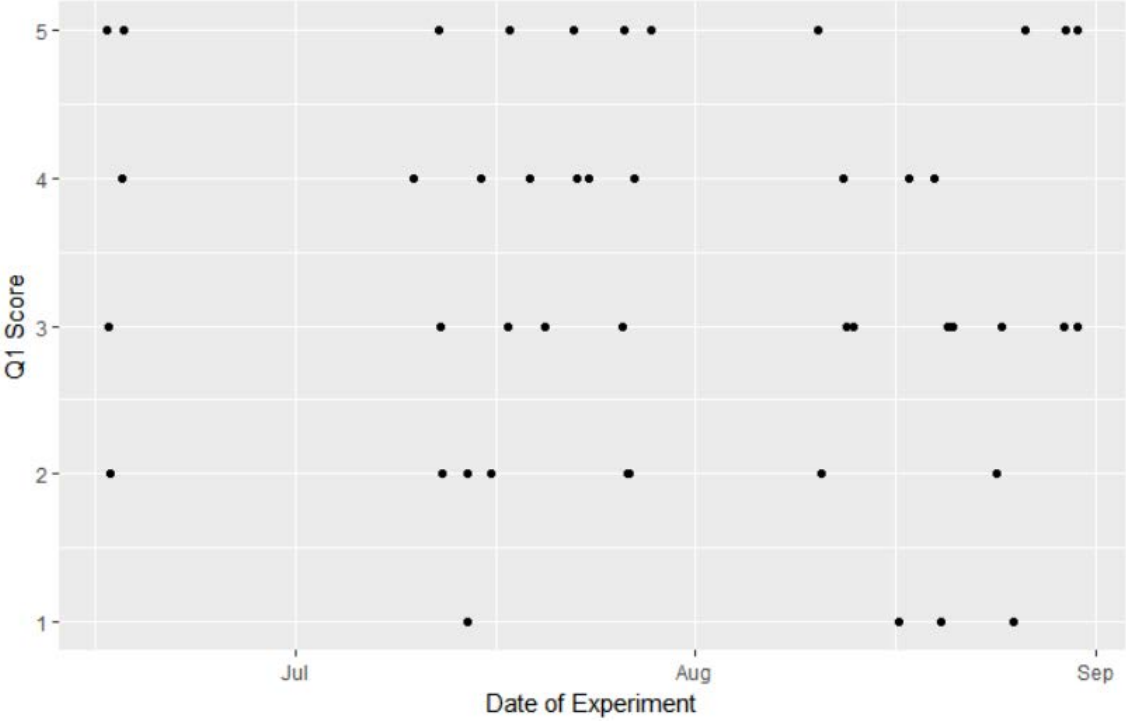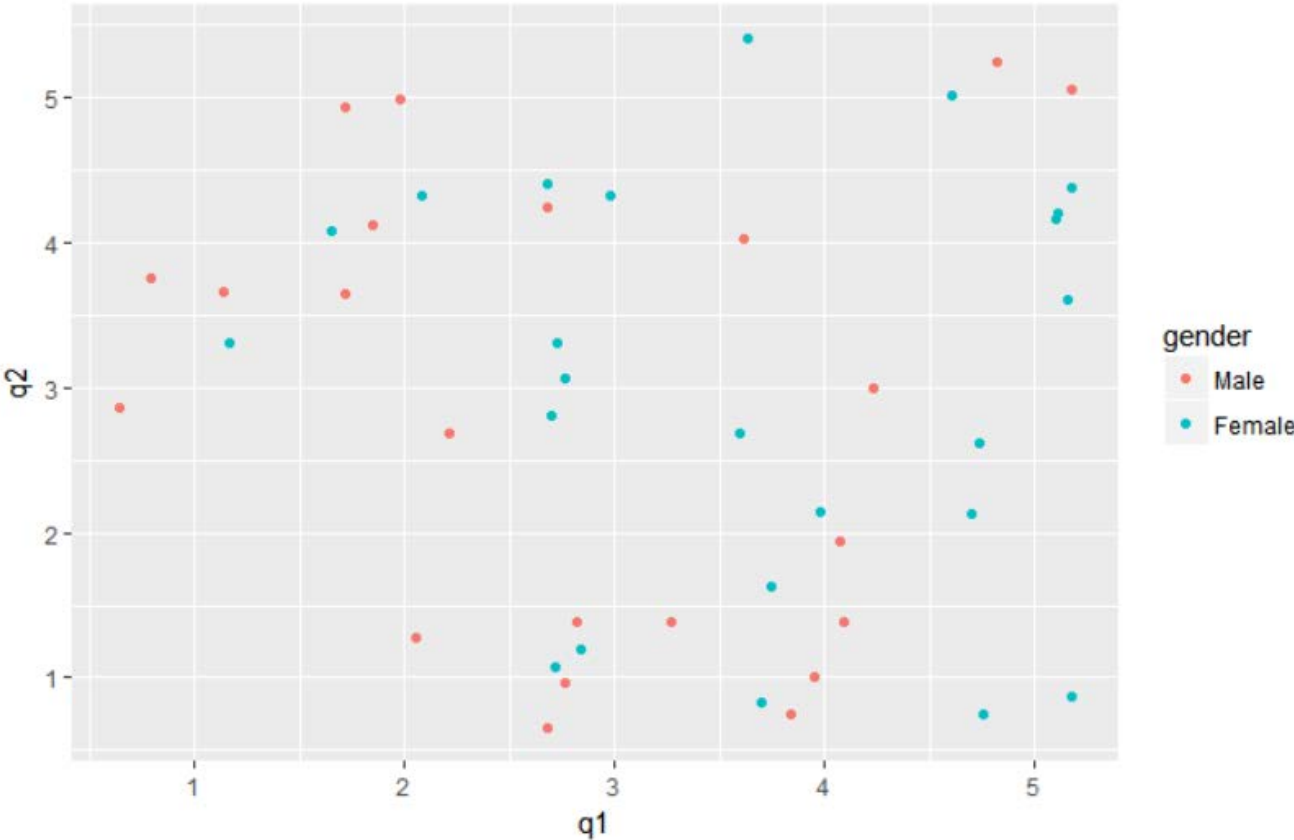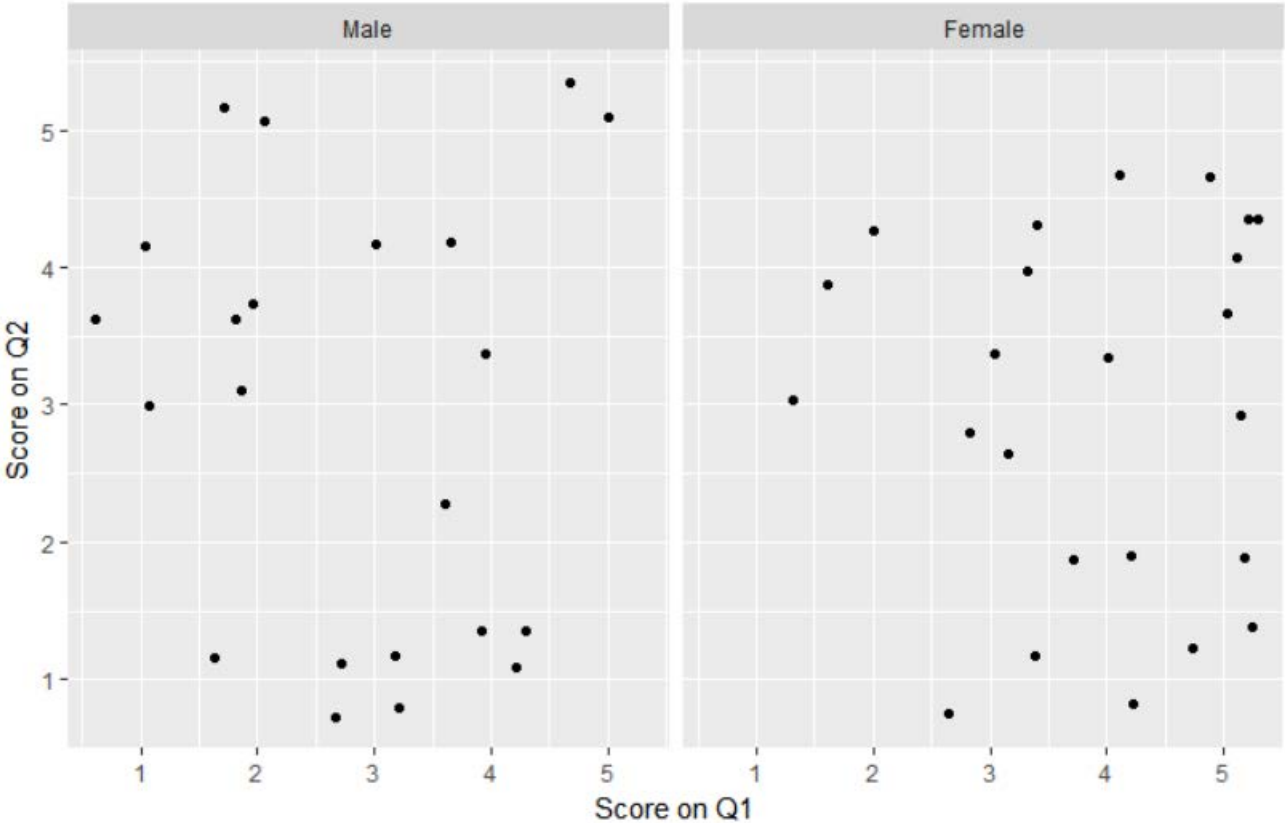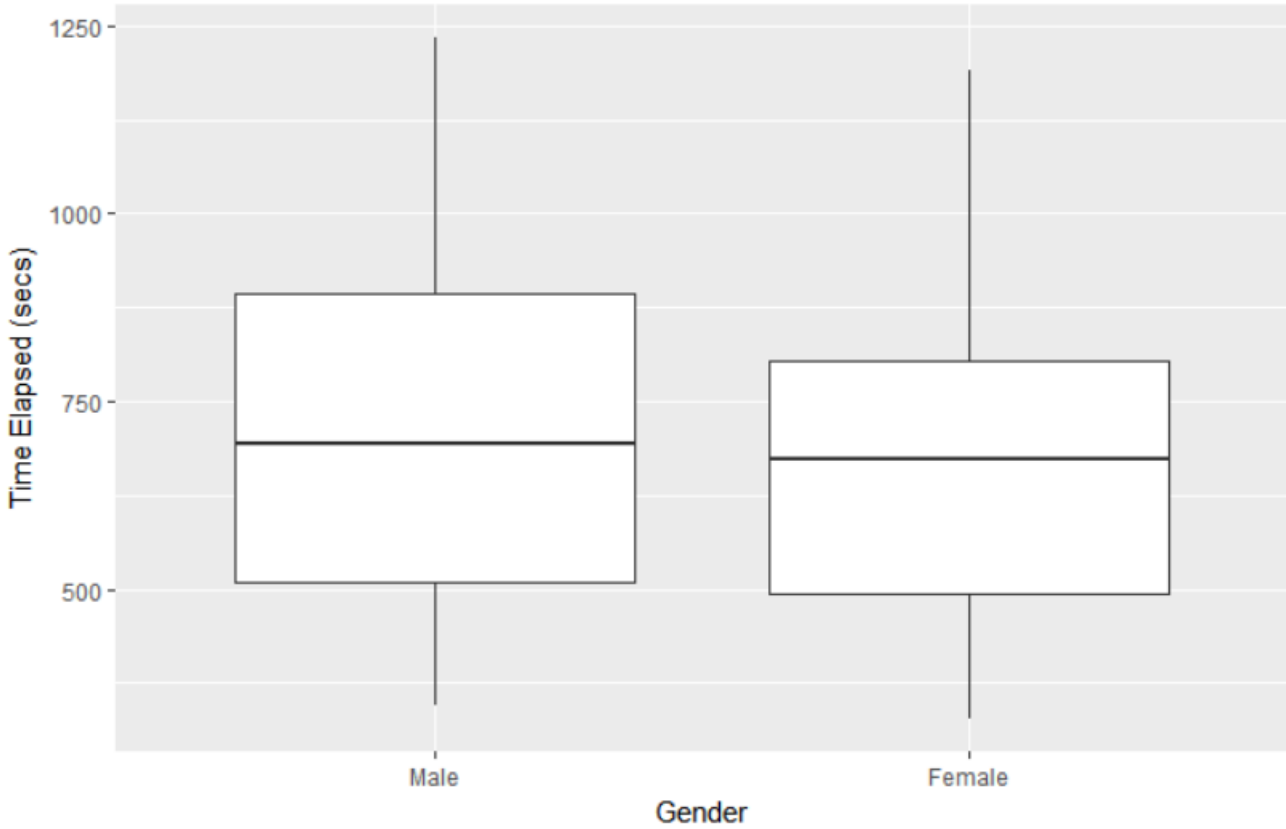
Fig 1

Fig 2

Fig 3

Fig 4

Fig 5