

Complete these tasks using RStudio. Discuss details with members of your assigned group, and record the answers you find, along with specific commands that assist you. If some sidelight question comes up, feel free to write about it and the modifications to commands that would help get at their answers.

For group discussions and recording answers, point your browser at this link to Etherpad

<https://pad.disroot.org/>

When prompted for a name of a "pad" to open, use the name

s145-31aug2021-gX

but replace the X with your group number. Use hyphens not underscores, and no capital letters. Make sure you get this right, or I will not be able to find your work. Use the "users" icon at the far right of the tool bar at the top to add your first name. Delete the initial text, beginning with "Welcome to Etherpad!" and concluding with a website. Replace this with a list that includes the full names of all users in this day's Group X.

The Groups:

Group 1: Beumer, Hoekema, Johnson, Miedema

Group 2: Brown, Fisher, E. Harsh, Shippy

Group 3: Arthur, Favila, LaCroix, Tanis

Group 4: J. Harsh, Knudson, Ngandu, VanderLugt

Group 5: Bieri, Lanser, Millen, Vroon

The tasks/questions:

1. Load a package called NHANES, whose primary purpose is to make available a data frame also named NHANES.

```
head(NHANES) # displays the first few rows
```

Determine how many cases are in the data frame, how many variables, what is the target population, and whether this can be considered a *simple random sample* from that population.

2. One of the variables in this data frame is named Poverty. Is it categorical or quantitative? How is it measured? What proportion of people in this data set live below the poverty line?
3. What is the largest number of pregnancies (see variable nPregnancies) among respondents?
4. One quantitative variable in this data frame is BMI. Does it work to create a frequency table on this variable? Do you see any drawbacks to using tally() on a quantitative variable? Does your observation apply to all quantitative variables? (How does it work out if you use it on nPregnancies, for example?)
5. Execute the command

```
gf_histogram(~ BMI, data=NHANES)
```

which produces a histogram. How is this like a bar chart? How does it differ from a bar chart? Can you describe in detail the process the software follows in order to produce this histogram from the BMI data?