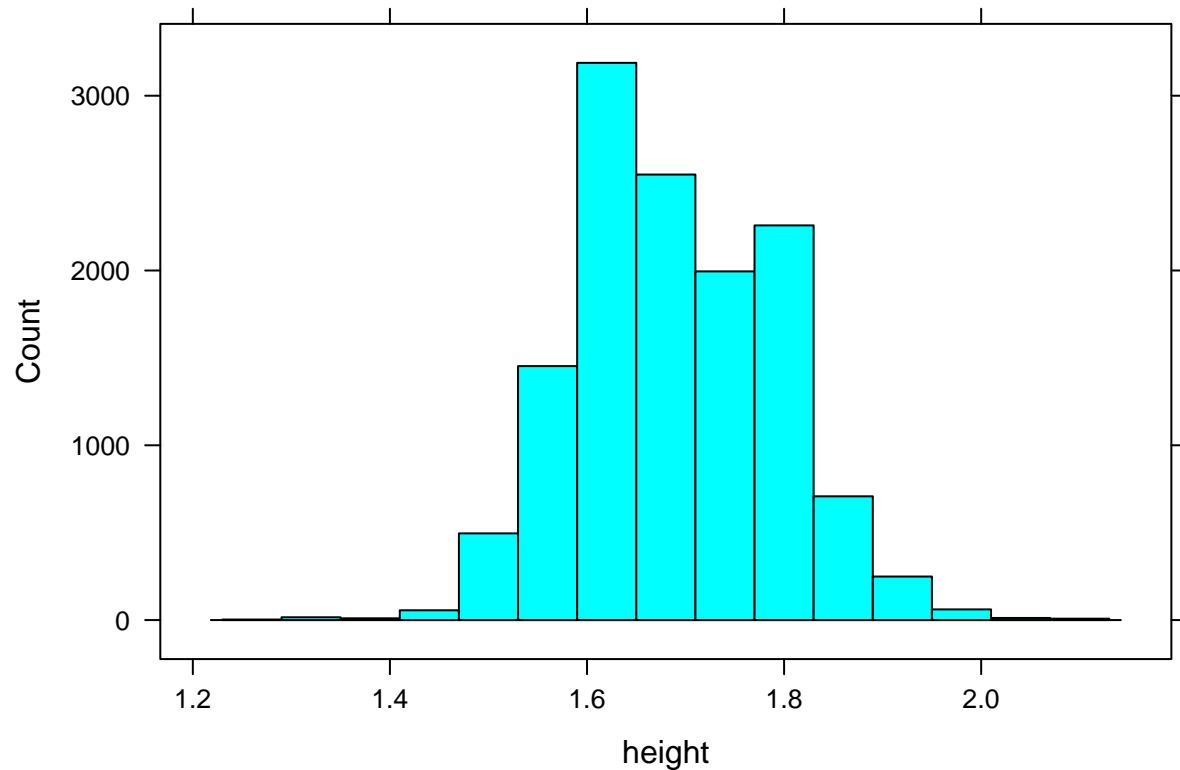# What's your status single?
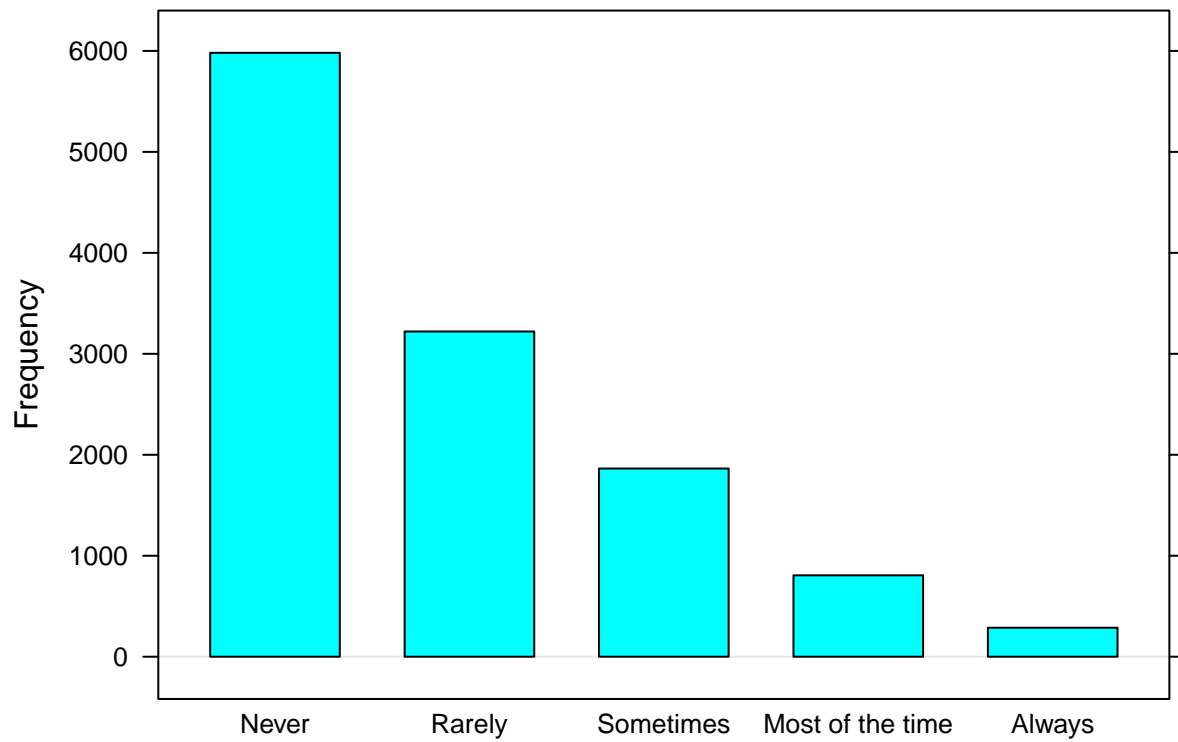
Directions: Follow along with the slides and answer the questions in **BOLDED** font in your journal.

Where'd we leave off . . .
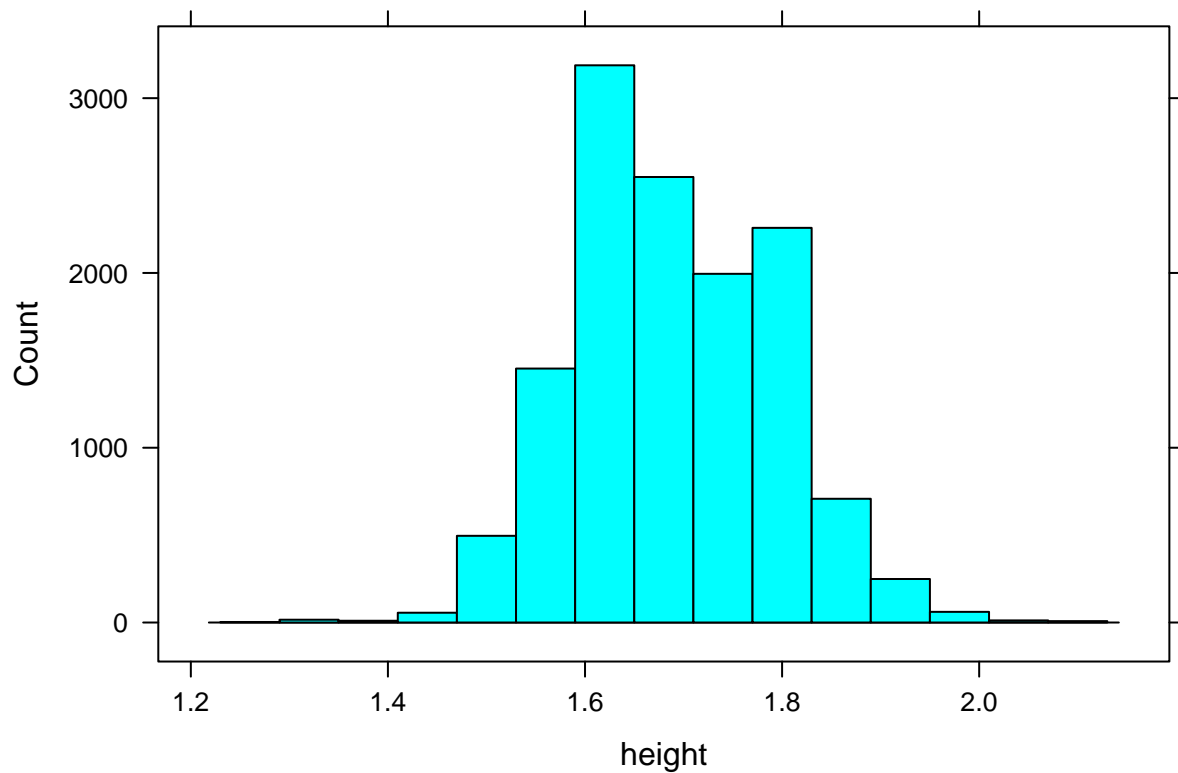
- The first lab ended. . .
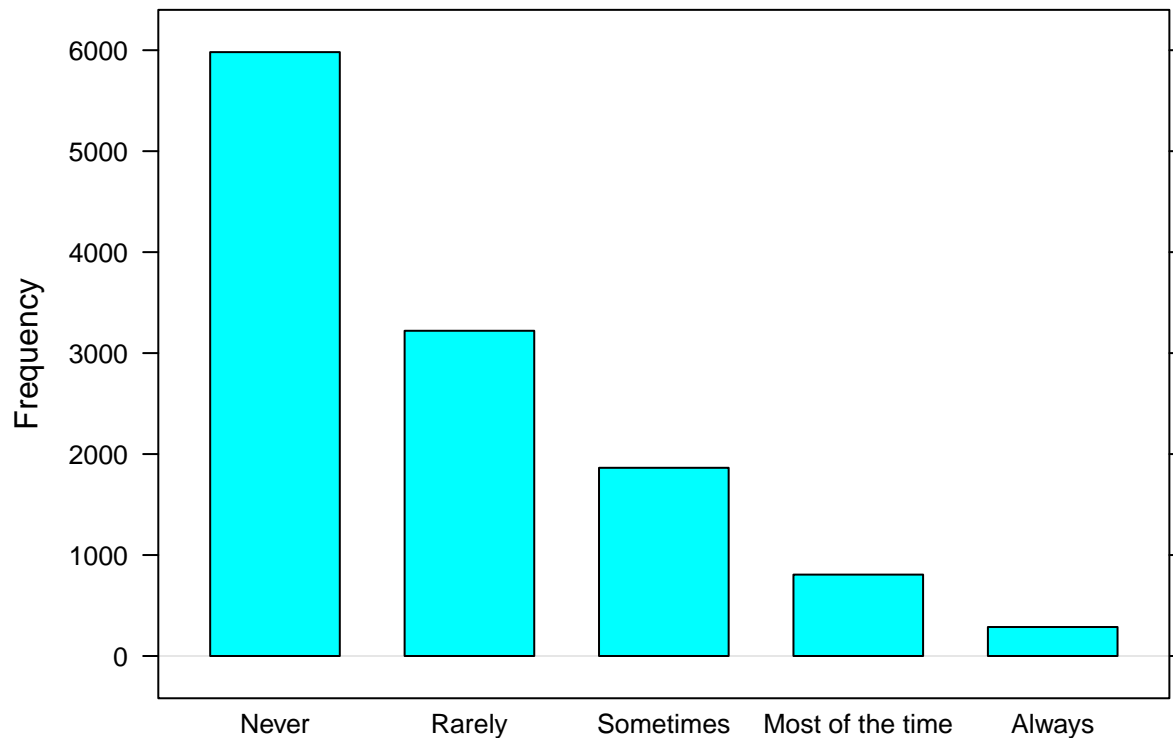

. . . with these two plots:

Where'd we leave off . . .

- This plot has an x-axis with numbers

- This plot has an x-axis with categories

Interesting. . .

- We know then that data comes in *at least* two different varieties.
- **Numerical variables** have values that are typically measured in units

  - Ex: Feet, inches, pounds, hours, sodas per day, etc.

- **Categorical variables** have values that describe what category the observation belongs to.

  - Ex: Heads or tails, red, green or blue.

Before we start!

- Load the CDC data from before by using the following command in the console.
  - Remember how to load this data set for future labs.

```
data(cdc)
```

Categorical Variables

- Have *values* that describe the *category* an observation belongs to

- For example, view your data by typing `View(cdc)` into the console.
  - Find the `gender` of the first person in the data.
  - Which *category* of `gender` does the person belong to?

Categorical Variables

- Now type the following command into the console to view the names of the variables.

```
names(cdc)
```

- **Write down 3 variables that you think are *categorical* variables and why you think that they're categories**
- **View your data and write down the different *values* (or categories) for each of the 3 variables you chose**

Bargraphs, a.k.a. Barplots

- *Bargraphs* are one of the best ways to *visualize* categorical variables.
  - One axis (x or y) will have the different categories.
  - The other axis will have the number of observations (or *Frequency*) that fall into each category.
- Type the following into the console to create a bargraph (Hit the *Zoom* button in the plot pane to make it larger).

```
bargraph(~helmet, data = cdc)
```

- **Explain what the values on the x and y axis mean. Which categories occured the most & the least often?**

More on Bargraphs

- Bargraphs are sometimes easier to read when the bars are horizontal.
  - Run the following command (Make sure to spell `TRUE` in all capital letters):

```
bargraph(~helmet, data = cdc,
         horizontal = TRUE)
```

- **In your opinion, are the vertical bars or horizontal bars easier to read for visualizing the `helmet` data. Why do you think that?**

More on Bargraphs

- We also sometimes want to *split* each bar in our plot by *grouping* them into seperate categories.
  - Run the following command to **group** the bars for each category based on each person's `gender`

```
bargraph(~helmet, data = cdc,
        groups = gender,
        horizontal = TRUE)
```

Your turn

- **Practice making AND interpretting bargraphs using all three of the categorical variables you chose earlier in the lab**
    - **Choose a different categorical variable if you previously chose `helmet`.**
- **Interpret each graph by explaining:**
    - **What the categories are**
    - **Which categories occured more or less often**
- **Be sure to point out any interesting discoveries you make**

Numerical Variables

- Have *values* that are *measured* in *units*.

- For example, view your data by typing `View(cdc)` into the console.
    - Find the `height` of the first person in the data.
    - How tall is this person? What do you think are the units?

Watch out!

- Sometimes variables that you think would be *numerical* are actually *categorical*.
    - You might think that `age` is *numerical* because it can be measured in *years*.
- View the values of people's ages in the CDC data by running `View(cdc)` again.
    - Since the values contain the words "years old", R inteprets these values to be categories!

Numerical Variables

- List the `names` of the variables again (You can go back to look up the command if you've forgotten)
    - **Write down 3 variables that you think should be *numerical* variables**
    - **List some possible units of measurements for each variable.**
- View your data
    - **Write down any variables you thought were numeric but were actually categorical (See the warning on the previous slide for help)**

Histograms!

- Just like how we used **bargraphs** to visualize *categorical variables*, histograms are useful for visualizing *numerical variables*
- Type the following to make a histogram for people's `height`

```
histogram(~height, data = cdc)
```

- **What do the values of the x-axis mean in terms of people's `heights`?**
- **What do the *widths* of the bins mean?**
- **What does the *height* of each bar represent?**

More on Histograms

- By changing the width of the *bins* of a histogram, we can change the amount of detail it shows.

    - Wide bins give us a very broad view of the data.
    - Narrow bins give us a very detailed view.

- A good histogram should strike a balance between being both wide & narrow

More on Histograms

- Run the following commands:

```
histogram(~weight, data = cdc,
          nint = 3)
```

```
histogram(~weight, data = cdc,
          nint = 30)
```

- **What would you say is the typical weight of these people? How do you think this compares to your class?**
- **Is it unusual to weight less than 65 kilograms? Why?**
- **Fill in the blanks: Most people weight between _____ and _____ kilograms. Explain how you chose your values.**