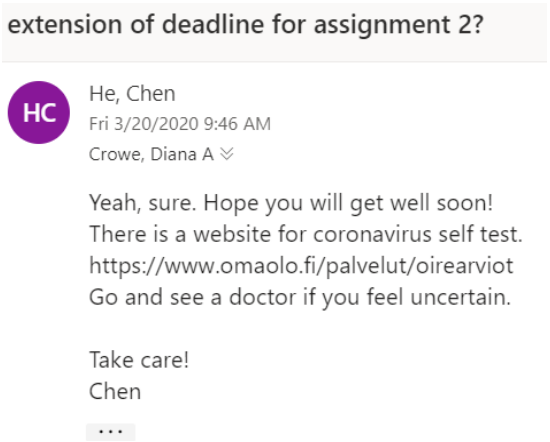
Assignment 2: Devising a static visualization

Thank you for allowing me as extension of the deadline due to illness! Me and my family are not well yet, but getting there and keeping isolated.



---

1. Download and pre-process the data.
2. Think about what question(s) you would like to answer using the data.
3. Transform / explore the data to answer your question(s) using some tools.
4. Apply a level of visual polish, such as adding legends, and changing typography.

Data Set:

File > download as... > Microsoft Excel (.xlsx) :

https://docs.google.com/spreadsheets/d/1vfRviO-nr1HFTfAEhtp2rK8SUfCqULA4sANQNR30g3M/edit?usp=sharing

The data “American community survey” contains the following:

1) **Age**: integer value ranging from 17 to 90
2) **Work class**: 8 nominal classes
3) **Education**: 16 nominal classes
4) **Level of Education**: integer value ranging 1 to 16
5) **Marital Status**: 7 nominal classes
6) **Occupation**: 15 nominal classes
7) **Relationship**: 6 nominal classes
8) **Race**: 5 nominal classes
9) **Sex**: binomial (male or female)
10) **Capital gain**: value in range 0 to 99999
11) **Capital loss**: value in range 0 to 2547
12) **Work hours per week**: value in range 1 to 99
13) **Native country**: 40 nominal classes
14) **Annual Income**: binomial: <=50k , >50k

Diana Crowe (student nr.: 012056152)

Questions to Investigate (starting point):

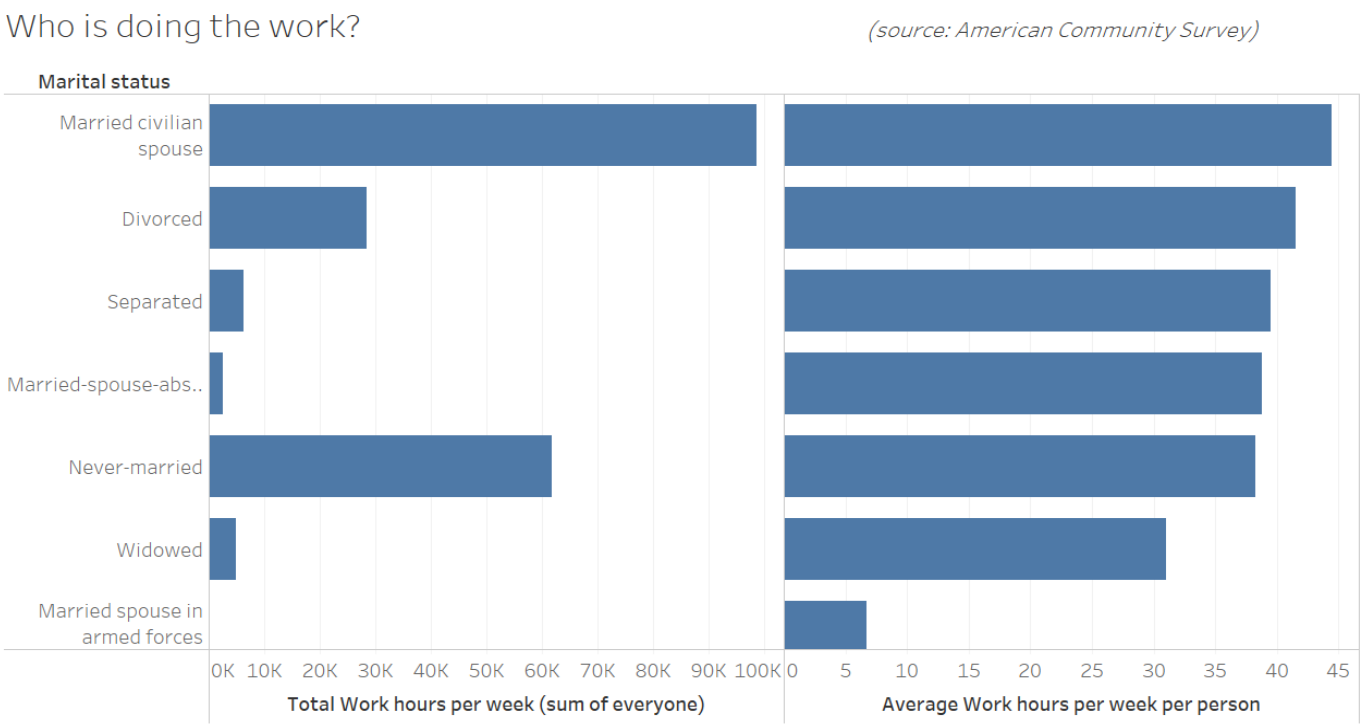
1) [Work hours per week] versus [[relationship] and [marital status]].

   Do people in a relationship/marriage work more or less than people not in a relationship/marriage?

2) Income and marriage: **Do husbands earn the same as wives?**

3) **Do people who work the same amount of hours/week and have the same occupation earn the same?**

Question 1 /Figure 1



I wanted to investigate how the marital status impacted on how much one had to work per week.

I decided on a double bar chart:

- Bar graphs are better for comparing larger changes or differences in data among groups (current case).
- It makes it visually easy to compare sets of data between different groups at a glance. In my chart we can directly compare the sum of the weekly hours worked by everyone in the same situation and the average weekly work hours for each person in each marital group (had to calculate that when pre-processing).
- The graph represents categories on one axis (marital status in this case) and a discrete value in the other (total hours worked per week for the category, and average hours worked per week for each individual in that category). The goal is to show the relationship between the two axes.
- Bar charts allow us to recognize patterns or trends far more easily than looking at a table of numerical data.
- I chose a horizontal bar chart for two reasons:
    - there is plenty of room for the long label along the vertical axis

    - I could have two bar charts in one figure so that we could directly compare the total work hours of each category and the average hours per week per person

- There is only one consistent colour filling the bars so at to not distract the viewer. The colour provides a good contrast between the figure in the foreground and the background.
- There are gaps between the bars of the bar chart so that it doesn't look like a histogram.
- No background grid (less is more). Grids are distracting and unnecessary when the goal is the bigger picture.
- I made the bars a bit fatter to separate the categories a bit more and make the chart more aesthetically appealing.
- In my chart, the amount for the average work hours per week were closer in value in the different groups, so I ordered those in descending order. This makes it intuitively easier to distinguish their relative amounts.
- Added a title and the source of the data.
- I did not use 3-D effect so as to not distract the viewer.
- The scale starts at zero.

Some interesting conclusions from looking at the chart:

- The bulk of the work is done by married people with civilian spouses, and they are also the people who work on average more hours per week. This is unexpected since one would assume that, if there are two people contributing to a household, they wouldn't need to work as hard to earn money*.
- The group that contributes the second largest amount of total hours worked is never-married people, but they tend to not need to work as hard – they are only in 5th place in the most hours worked per week.
- Divorced and separated people have it much worse than never-married people since they come in second and third place respectively place for most hours worked per week (on average).
- Widowed people work short weeks.
- People married with spouses in armed forces don't work much (in paid jobs).

Question 2 /Figure 2

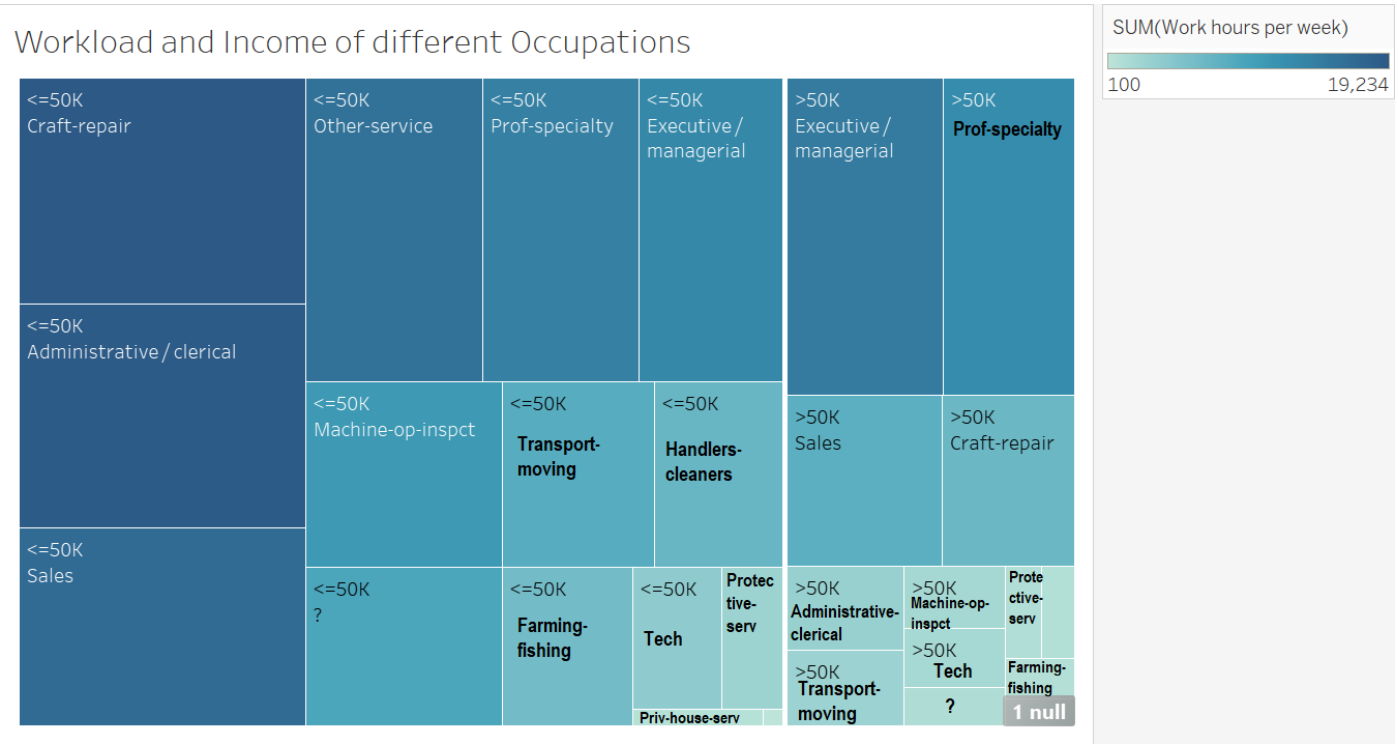
	<=50K	**>50K**
husband	1081	919
wife	115	130

After the first crunching of numbers I decided that there wasn't enough data to make any interesting visualizations comparing the income from husbands and wives. It is obvious that there is not enough data in our spreadsheet: there is a total of 2000 husbands but only 245 wives. We would wish to directly compare the income of both married people.

* Incidentally, this sheds a bit of light on the results from Figure 1. The vast majority of the people married with a civilian spouse are husbands, so we could hypothesize that maybe the reason they need to work so many hours per week is that several might be the sole wage-earners of the household.

Diana Crowe (student nr.: 012056152)

Question 3 /Figure 3



I made a treemap to explore whether people in the same occupation who worked around the same amount of hours earned the same. It is suitable enough for a qualitative quick study, though it can be tricky to judge the difference between two shades.

Some interesting findings:

- For around the same number of hours worked, some people earn a lot more than others in the following fields: Prof-specialty, Technical, Protective Services, and Executive/ Managerial
- There is a divide in several occupations where a handful of people earn more than their peers: Craft-repair, Administrative-clerical, Machine-op-inspct, Farming-fishing, and Sales
- Occupations where the earnings are consistent for everyone include: Private-house-servants, Handlers-cleaners, and Armed Forces (all earning under 50K per year)

Ultimately, I am not happy with this visualization. If there would be more detailed income data instead of just <=50K and >50K, we could make much more interesting and useful graphs since we could then plot both the work hours and the income. The treemap was the best I could come up with (so far) for the current data, but it is not sufficiently easy to glean details from it…

Diana Crowe (student nr.: 012056152)