

Data Science

Survival Skills

Homework 4

Homework 4

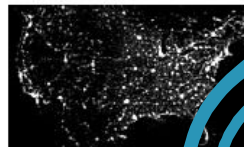
In this week's lecture, we covered data exploration and visualization.

In this homework you will work with the Census Income Dataset.

Census Income Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Predict whether income exceeds \$50K/yr based on census data. Also known as "Adult" dataset.

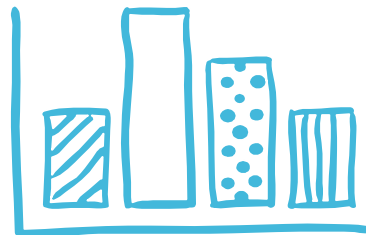


Data Set Characteristics:	Multivariate	Number of Instances:	48842	Area:	Social
Attribute Characteristics:	Categorical, Integer	Number of Attributes:	14	Date Donated	1996-05-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	722141



Homework 4: Tasks 1/2

- Load the **Census Income** dataset we provided via StudOn.
- Create **three** suitable plots to visualize this data:
 - a) Age distribution of respondents
 - b) How often does each relationship status occur?
 - c) Investigate and visualize how many respondents have a salary of $\leq 50k$ or $> 50k$ within each educational level. (Hint: You can group the values, i.e. you can count “ $\leq 50k$ ” as value 0, “ $> 50k$ ” as value 1)
- Each of the three plots must be a **different** type of plot.
- Make sure that the font size is large enough!



Homework 4: Tasks 2/2

- Save the plots you created as vector graphics (.svg)
- Create a figure in Inkscape and insert your three plots
→ <https://inkscape.org/de/>
- Save the figure as SVG file
- Export it as PNG file for your slide



- **Upload to StudOn:** The SVG file
- **Slide:** The figure you created
- **Slide:** A caption that describes why each plot is an adequate visual, i.e. explain your thought process that went into the plot.



Homework 4: Example

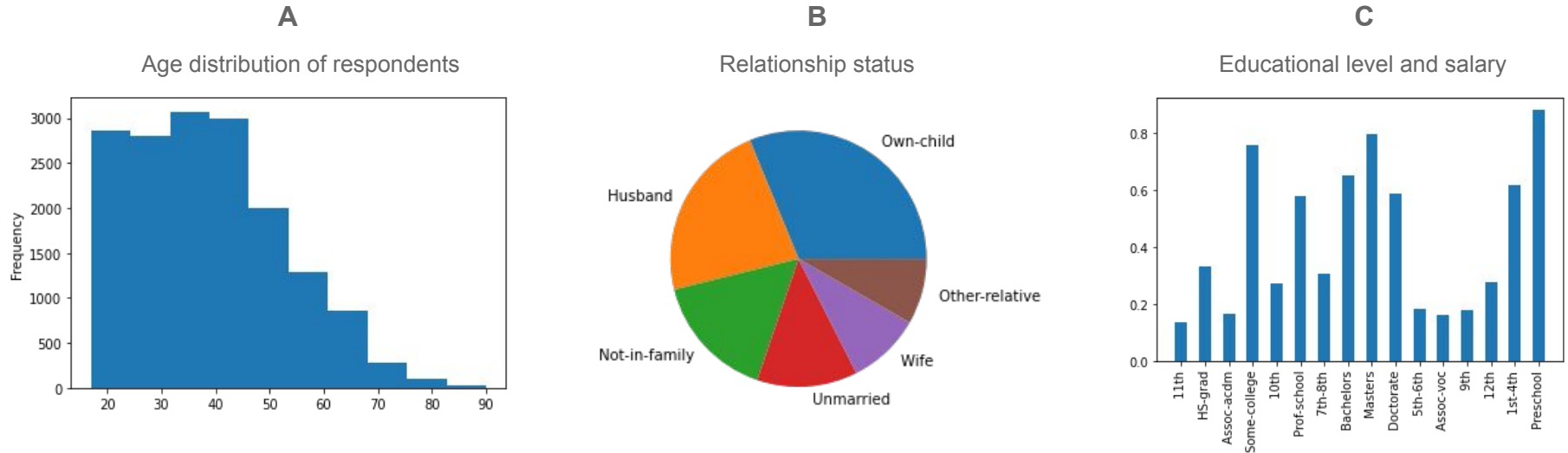


Figure 1: Visualization of the Census Income dataset. (A) This plot is an adequate visual for representing the age distribution in a effective way due to some parameters. Be creative! (B) This plot is an adequate visual because [...] (C) This plot is an adequate visual because [...]

Homework 4: Example

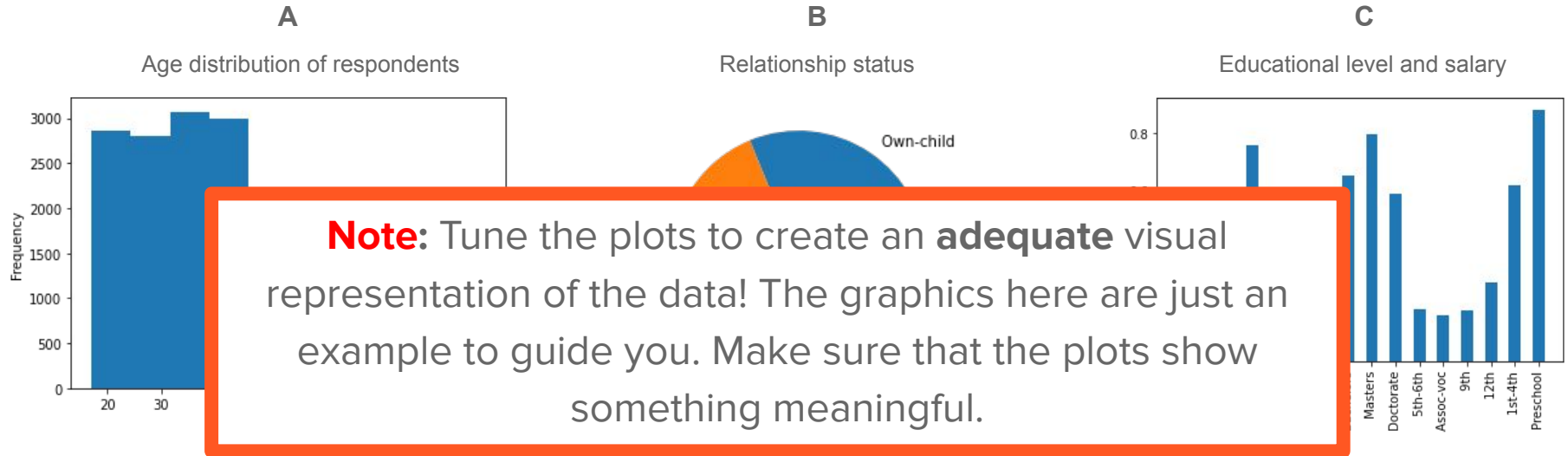


Figure 1: Visualization of the Census Income dataset. (A) This plot is an adequate visual for representing the age distribution in a effective way due to some parameters. Be creative! (B) This plot is an adequate visual because [...] (C) This plot is an adequate visual because [...]

Homework: Requirements

You must complete **all** homework assignments (**unless otherwise specified**) following these guidelines:

- **One** slide/page.
- **PDF** file format only.
- It has to contain your **name** and **student (matriculation) number** and **IdM** in the down-left corner.
- Font: **Arial**, Font-size: **> 10 Pt.**
- Answer **all** the questions and solve all the tasks requested.
- Be careful with **plagiarism**. Repeated solutions will not be accepted!