

Data Science Survival Skills

Homework 3

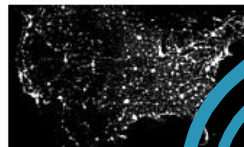
Homework 3

In this week's lecture, we covered data exploration, statistics, distributions and some visualization basics. In this exercise 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 3: Tasks 1/2

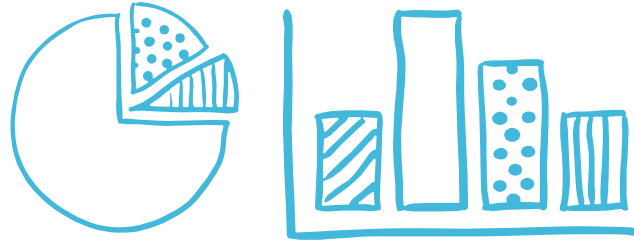
- Create three different plots (plot types) using the **Census Income** dataset we provided via StudOn. The following data should be presented in a proper way:
 - a. Age distribution of respondents.
 - b. How often does each relationship status occur.
 - c. Investigate and visualize if there is a correlation between educational level and salary (Note: you can see ' $\leq 50K$ ' as value 0; ' $>50K$ ' as value 1).
 - Shortly explain why your figure can be an adequate visual, i.e., explain your thought process that went into that figure.
- **Slide:** Each plot a), b) and c)
- → **Slide:** Your answer



Homework 3: Tasks 2/2

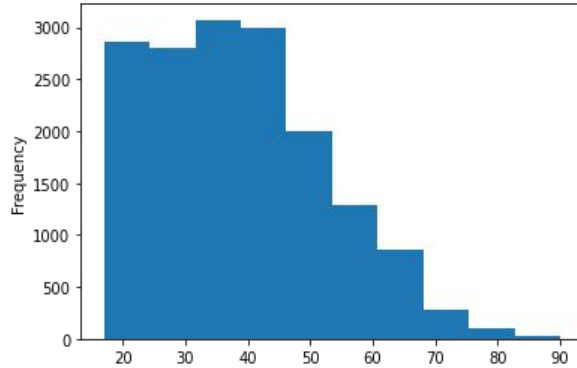
- Perform some reasonable statistics using the available data and write the results in a sentence. You can't repeat the example made by us.

→ **Slide:** E.g., “on average, a male united states American labors 42 hours per week, whereas a female works 36.”



Homework 3: Example

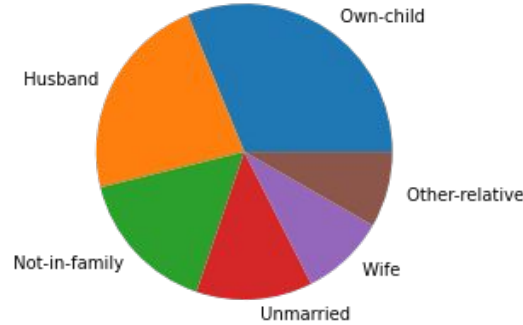
a) Age distribution of respondents.



This is an adequate visual because it represents the data in a effective way due to some parameters. Be creative!

- List some points
- List some points

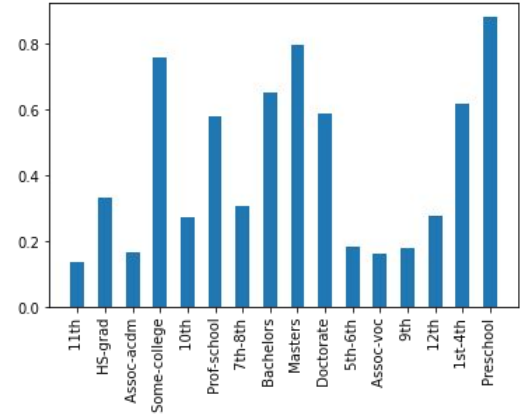
b) Relationship status.



This is an adequate visual because it represents the data in a effective way due to some parameters. Be creative!

- List some points
- List some points
- List some points

c) Educational level and salary.



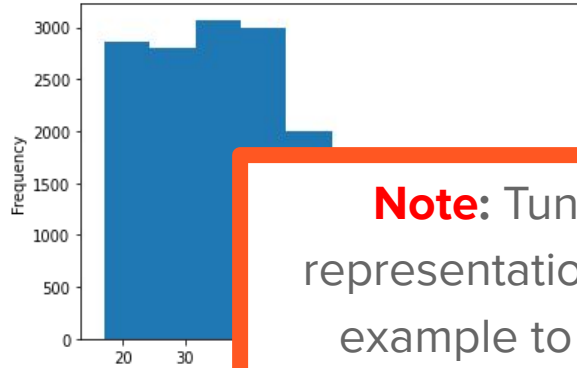
This is an adequate visual because it represents the data in a effective way due to some parameters. Be creative!

- List some points
- List some points
- List some points
- List some points

On average, a male united states American labors 42 hours per week, whereas a female works 36.

Homework 3: Example

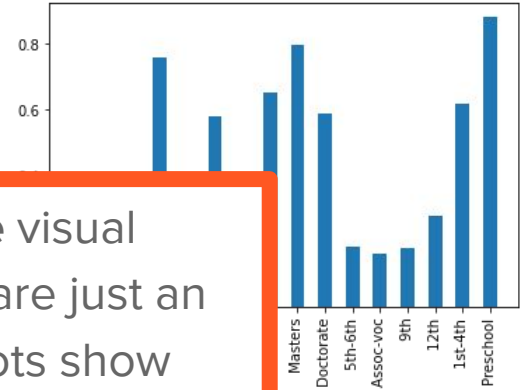
a) Age distribution of respondents.



b) Relationship status.



c) Educational level and salary.



Note: Tune the plots to create an **adequate** visual representation of the data! The graphics here are just an example to guide you. Make sure that the plots show something meaningful.

This is an adequate visual because it represents the data in an effective way due to some parameters. Be creative!

- List some points
- List some points

This is an adequate visual because it represents the data in an effective way due to some parameters. Be creative!

- List some points
- List some points
- List some points

This is an adequate visual because it represents the data in an effective way due to some parameters. Be creative!

- List some points
- List some points
- List some points
- List some points

On average, a male United States American labors 42 hours per week, whereas a female works 36.

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** 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!

And we are done!

Thank you