What are the factors deciding income over 50K?

Phillip Hyunho An, Christina Chou, Steve Kim, Stella Weng

36-315 Statistical Graphics and Visualization

Professor Matey Neykov [Department of Statistics]
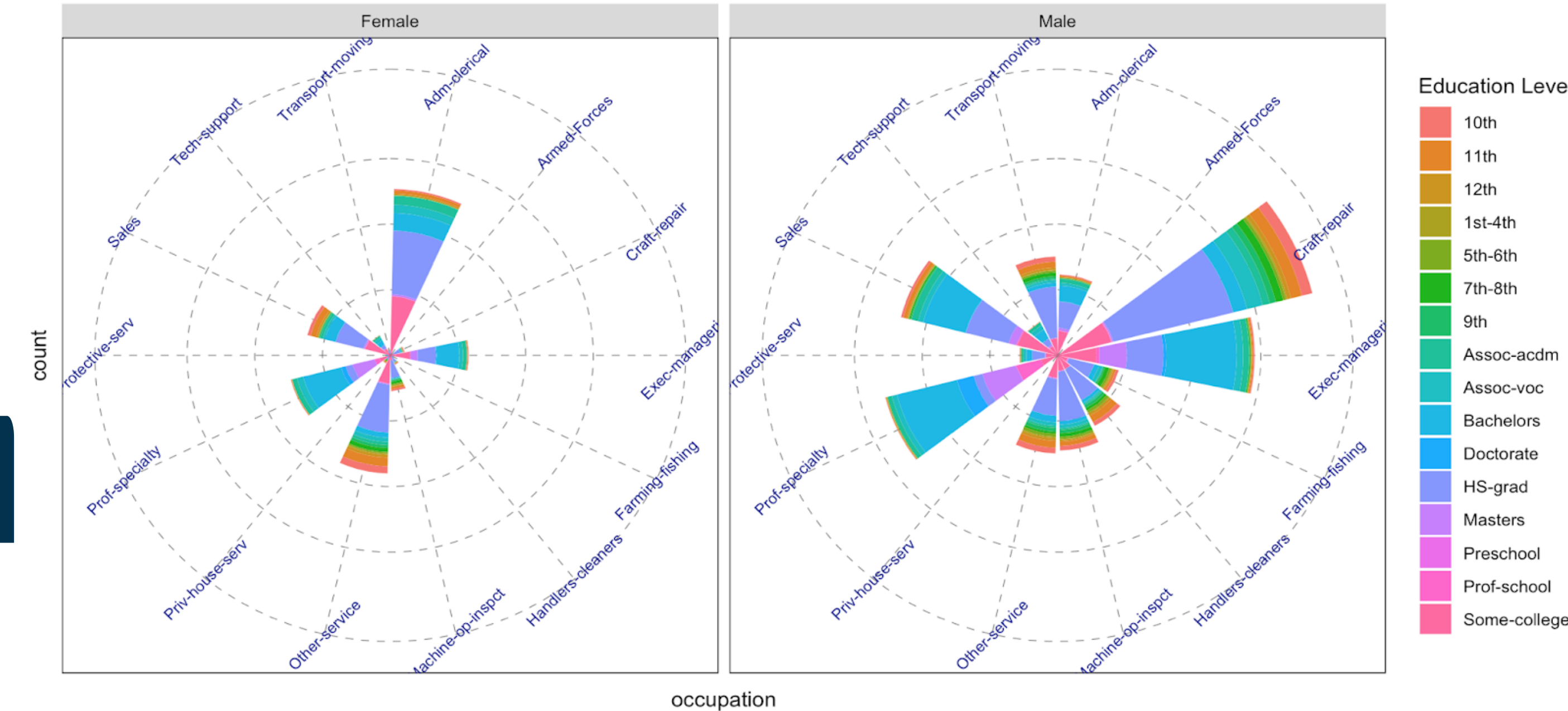
OVERVIEW OF DATASET

1994 Adult Data Set

- Predict whether income exceeds $50K/yr based on census data. Also known as "Census Income" dataset.

- Prediction task is to determine whether a person makes over 50K a year.

- Extraction was done by Barry from the 1994 Census database.

WHAT DO WE WANT TO KNOW?

- Will marital status influence annual earnings?

- How are continuous variables such as age, hour of working, education related?

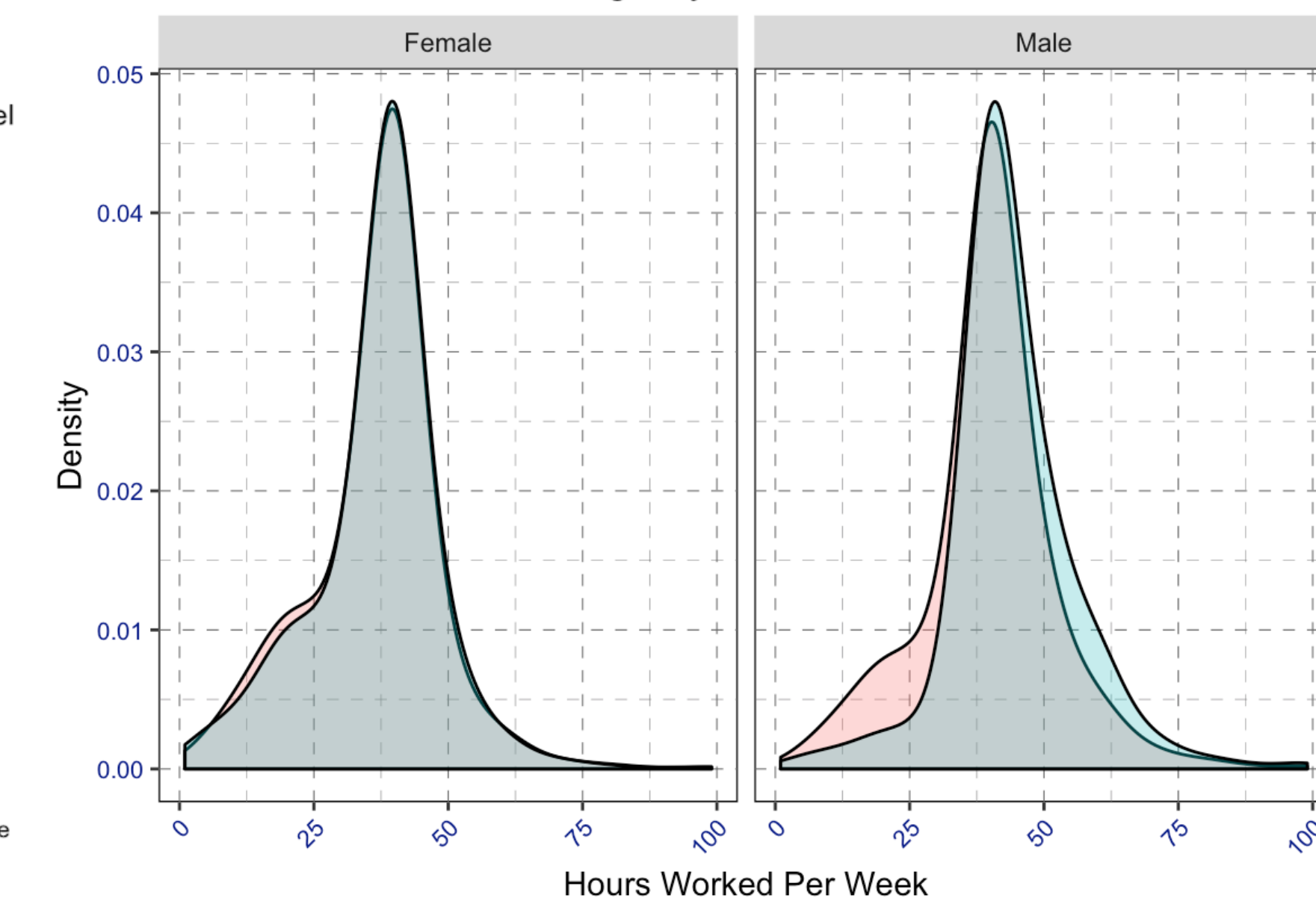- Which occupation is the most popular occupation?

MARITAL STATUS VS. INCOME

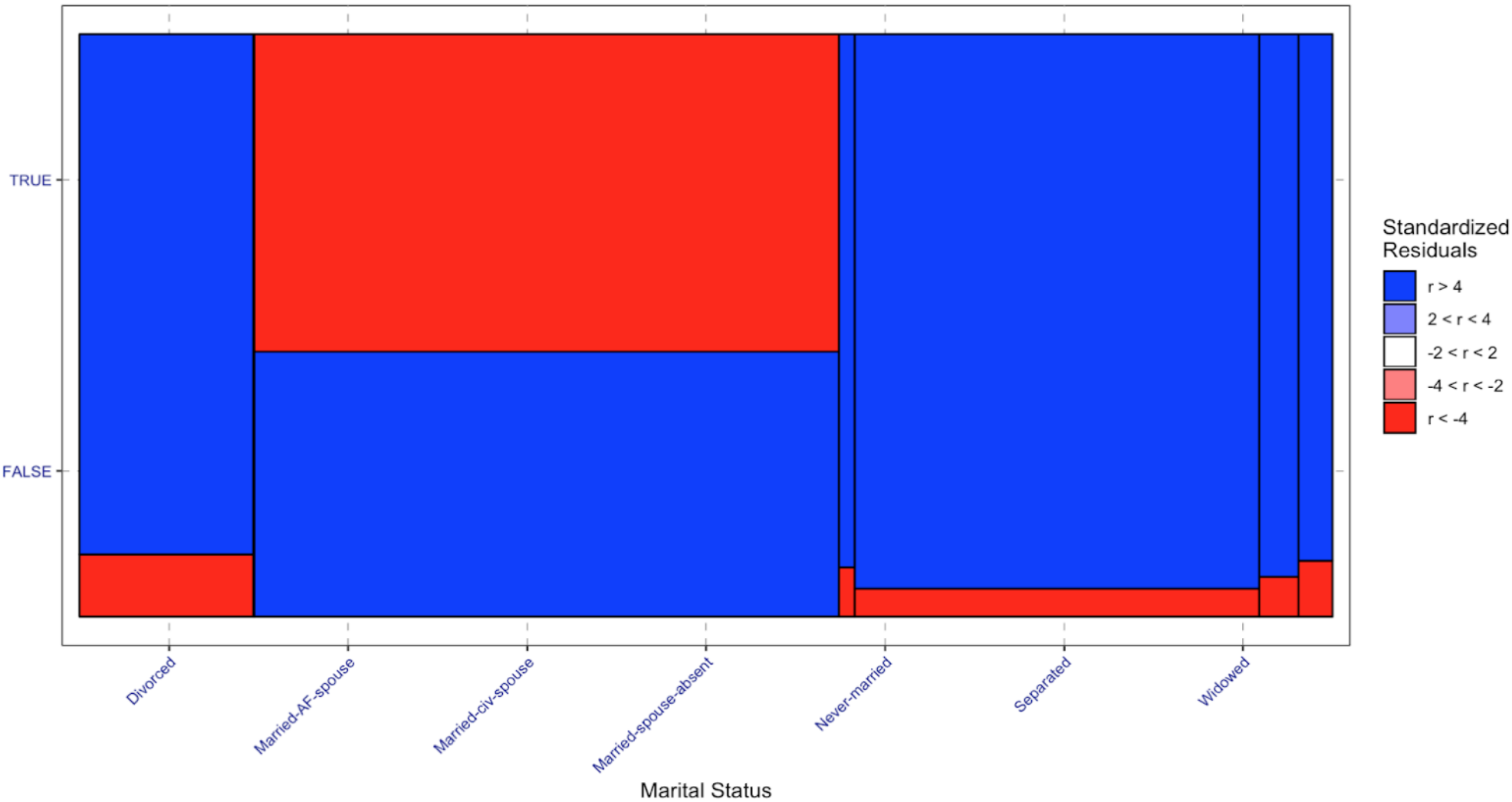Popular Occupation and Occupying Degree Facetted on Gender

- Expected education level for craft-repair is high-school graduate, but bachelors for executive-managing; required education level depends on the occupation.

Hours Worked on Average by Marital Status and Gender

- Density plot shows that there was not a major difference in terms of number of hours worked between men and women, with the men having slightly higher density at the higher number of hours.
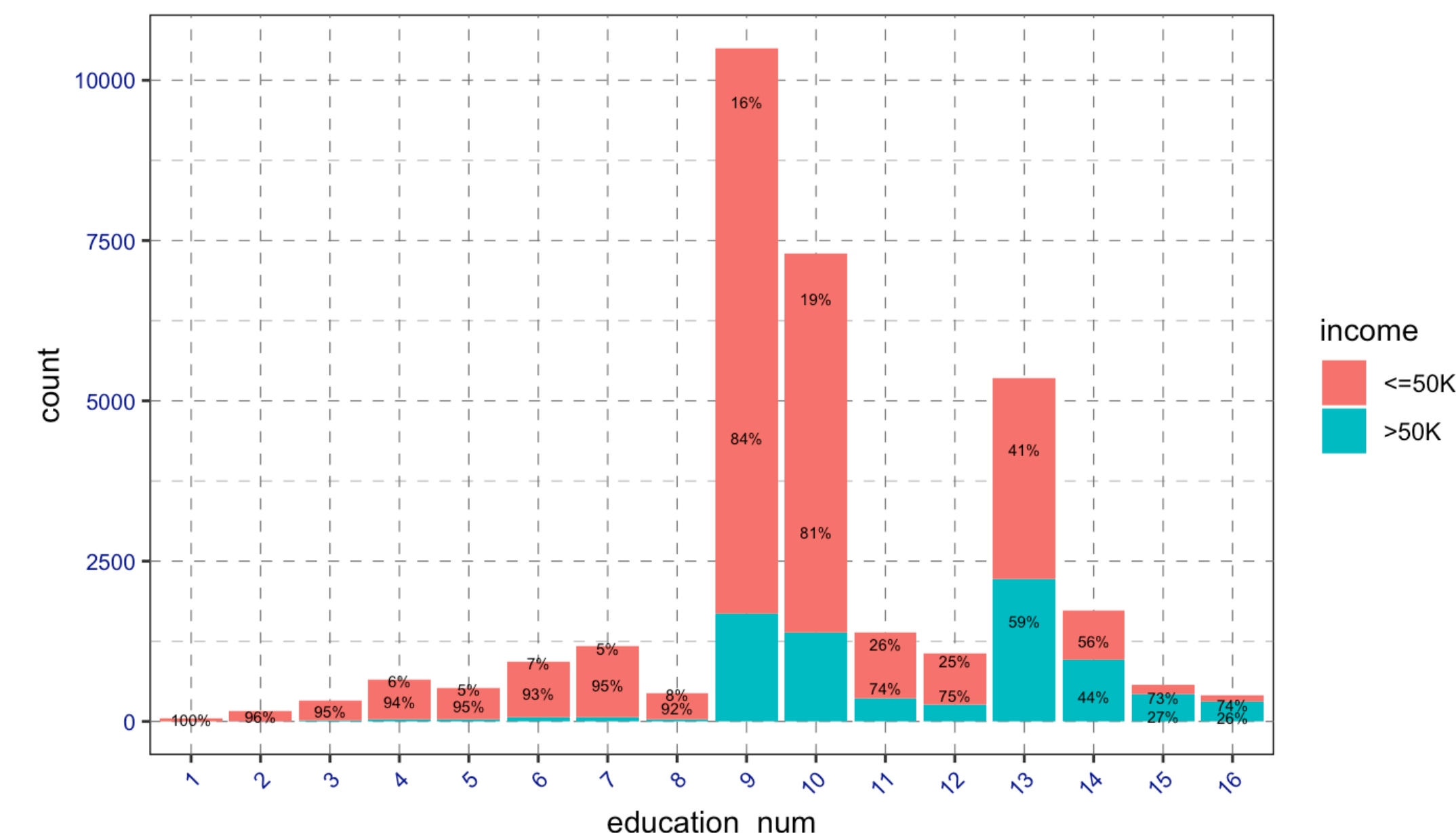

Relationship Between Marital Status and Earnings

- Unfortunately, people are much more likely to earn more than 50k when their marital status is divorced or married-spouse-absent and etc, but only 50 to 50 chance to get more than 50k when they are married.
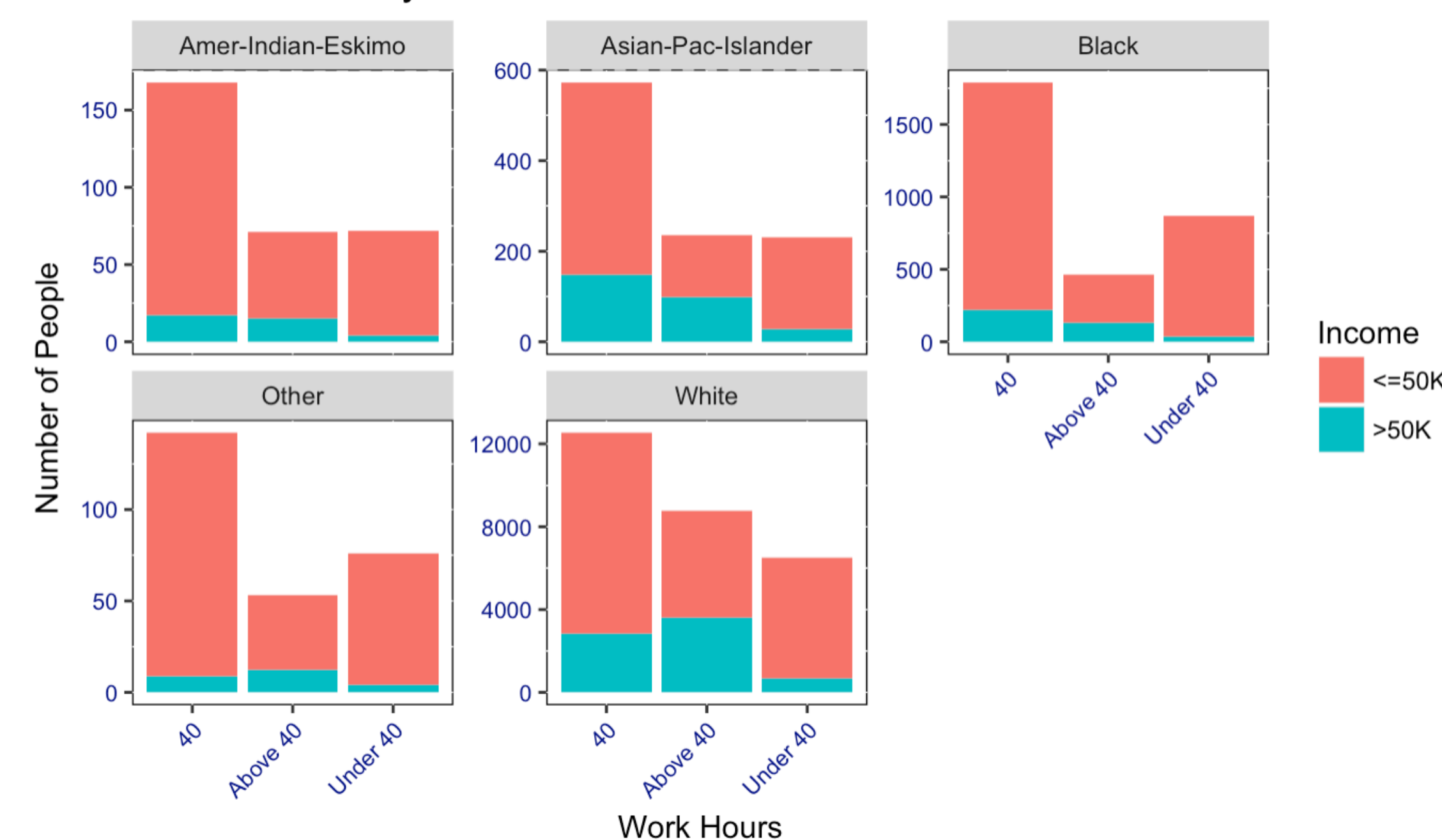
YEARS OF EDUCATION

Income Level with Years of Education

- As the years of education increases, the in group proportion of making over 50k a year also increased.

- For individuals who didn't have any forms of college education, less than 10% had an annual income greater than 50k.
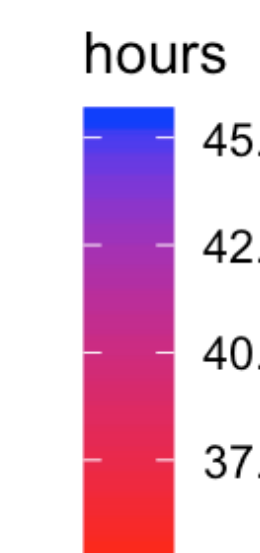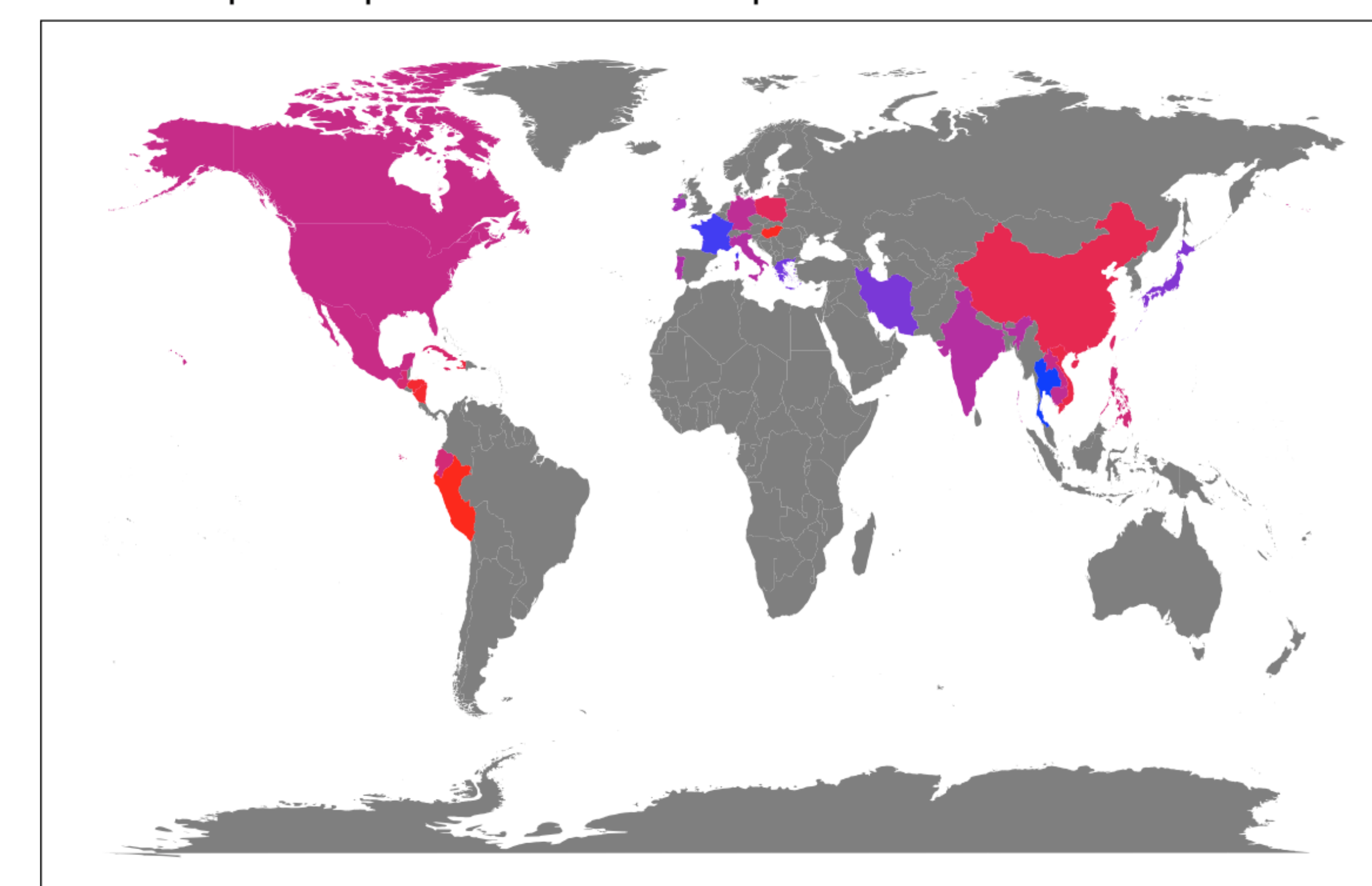
WORK HOURS AND RACE

Income Level by Work Hours and Race

- Based on the data, people who work above 40 hours per week have the highest possibility of earning more than 50k a year.

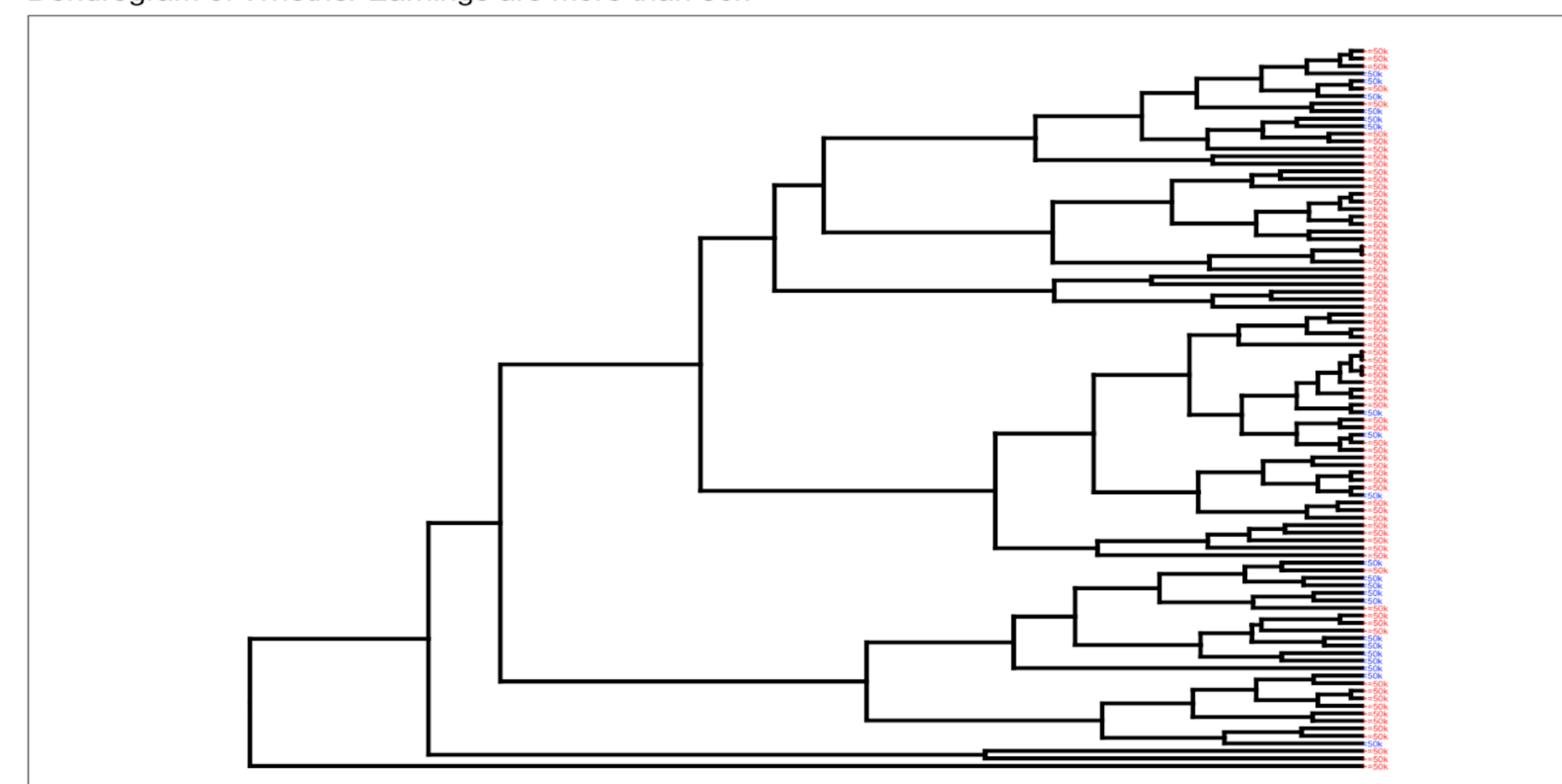## NUMBER OF HOURS WORKED BY REGIONS

World Map of Expected Work Hours per Week

- Demographic map showing the number of hours worked reveals that countries in Europe actually worked more on average; in contrast, China had very low average worked hours.
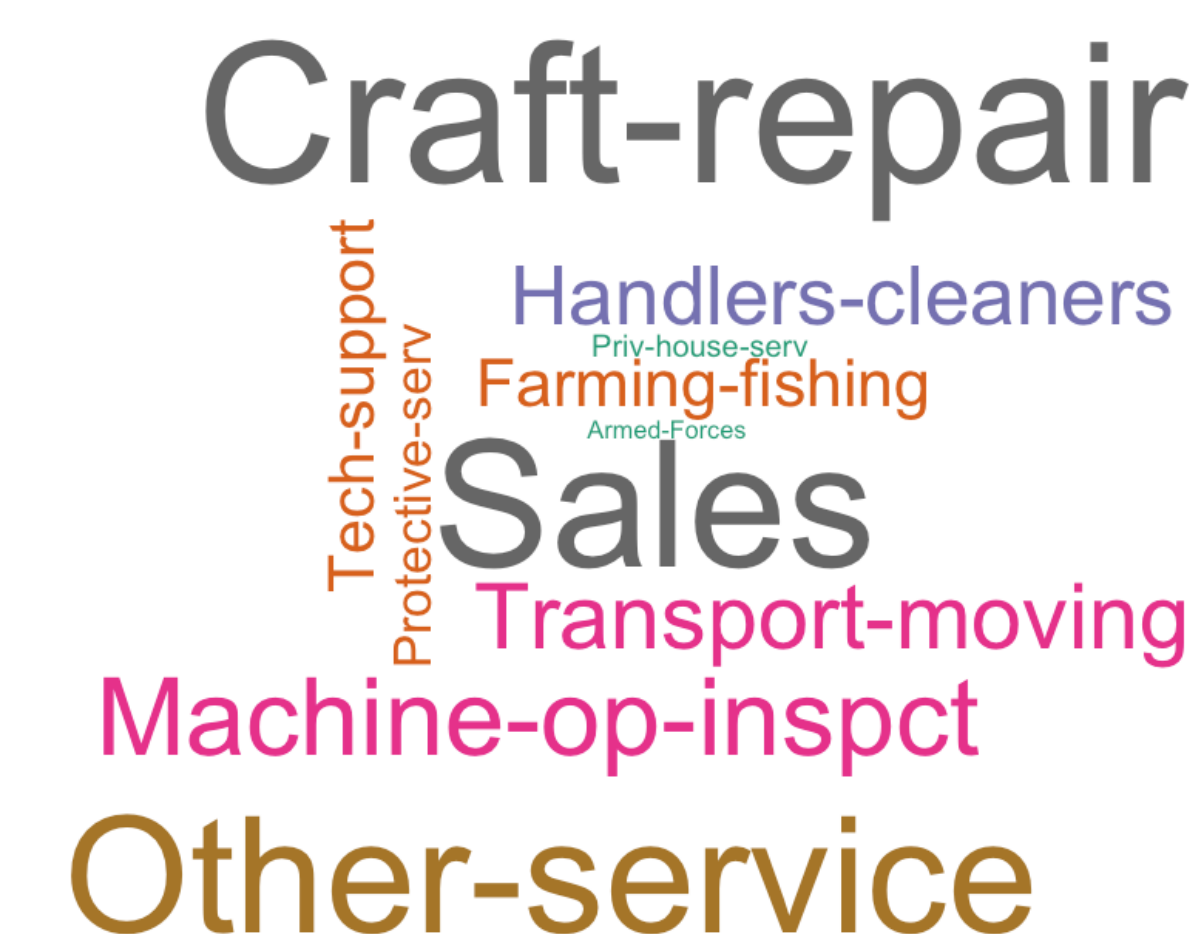
DENDROGRAM OF OCCUPATION

Dendrogram of Whether Earnings are more than 50k

- A dendrogram on earnings (whether>=50k) based on age, working_hours, and education level.

- There are lots of groups in different sizes; we can easily see individuals from the same group that are clustered together.

WORD CLOUD (OCCUPATION)

- Most popular occupation that earned more than $50k are in order of sales, craft-repair, and other-service.

RESULTS & CONCLUSION

- Overall, it seems that all of the major factors, such as age, gender, work hour, marriage status and education are related in some way to whether people earn more or less than $50,000 per year. However, we should take into account that this data set is from 1994 and that the relationships may have changed now. Also, there may be underlying interaction effects that we have not yet investigated.

- Divorced people who received more education and work more than 40 hours per week are more likely to earn more than $50k as an income. On average, North America has a higher mean weekly working hours while white individuals are more likely to receive more than $50k annual income across different working hours.