Final Group Work – Analysis of ANNUAL BASE SALARY

Group 4:

2021-09-21

**Structure of our analysis:**

1. Introduction
2. Executive Summary
3. Exploratory Data Analysis (Correlation)
4. Additional Data Analysis (Hypothesis testing, Regression Models)

**Executive Summary**

*(copied from Lady’s website)*

* 24k respondents, mostly female and from the United States
* Computing/Tech, education nonprofits, and government, health care were top 5 represented industries
* In the US salaries reported weremuch higher than the national average (HINT Compare EDA to other online data), reflecting a sample of mainly educated professionals
* On average, US men reported higher salaries than women, who reported higher salaries than non-binary people
* Asian/Asian-American reported the US highest salaries
* Find the top paid industries
* Find about bonus
* Years of field experience and education pays off disproportionally

**Cleaning the data:**

(What Wei did, and picture sent into the what’s app chat)

**EDA:**

* Who are survey respondents:
  + Gender: Mostly female
  + Race
  + Industry
  + Country
  + Age
* Salary analysis US only (as numbers of other countries Canada 1.5k, UK 1.3k, others more or less insignificant)
  + Age
  + US individual base salary compared to national household salary
  + Men vs. women vs. non binary
  + Salary by race
  + Base salary diff by industry
  + Age vs salary
  + Years of field experience vs salary 🡪 gradually higher
  + Education vs salary
  + Bonus

??? Do we want to include analysis from the other countries and compare to the US?? 🡪 but then we need to watch out for currency changes etc -.-

**Additional Data Analysis (Hypothesis testing, Regression Models, etc.)**

*Lots of factors at play I think! Agree that this effect is probably less obvious when controlling for location, education, sector, etc. I could easily be convinced that any or all of the following were true:*

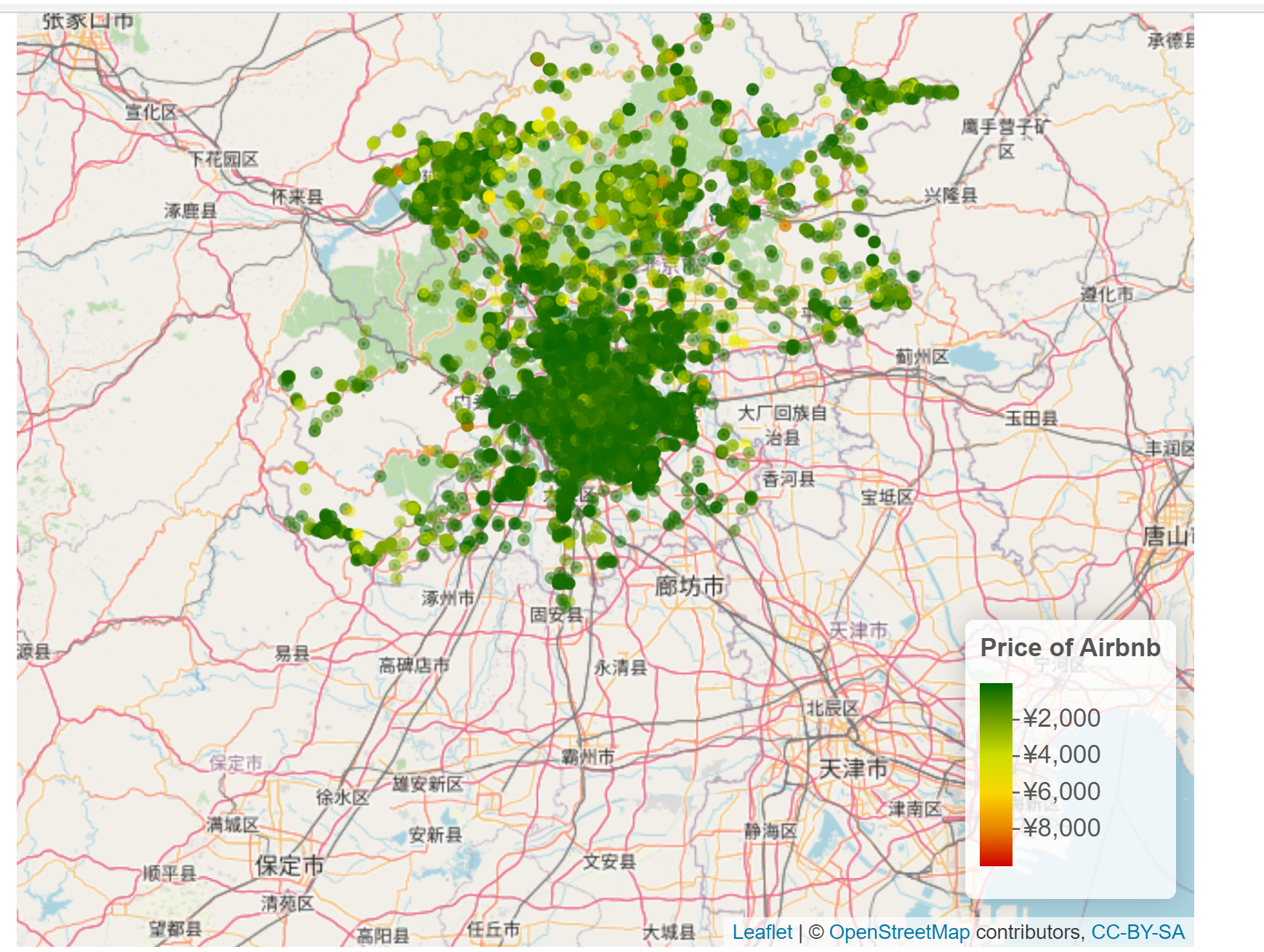
1. *Asian respondents more likely to work in tech (hypothesis!)  
   b) Asian respondents more likely to live in HCOL areas (LA, SF, Seattle all have pretty sizable Asian population and are also $$ to live in) (hypothesis 2)  
   c) Asian respondents more likely to have high level of education (hypothesis 3)*

* *If all above (1-3 hypothesis are true, we can explain better why Asians have highest salaries)*

*!!!CHALLENGE:*

If we decide to do go on a city level we could create a world map showing in intervals where people earn the most, it should show probably that in cities ppl earn more

* Show hotspots for bigger cities!!!

Sth like this but for salary:

* Convert all currencies to USD, then compare the mean and medians of the countries
* Same as above for bonus compensation
* seeing the same graphs with Medians instead of Means
* bunch of fascinating histograms (of what???)
* At what age to ppl earn the most, is it more with age or is there a clear peak?
* I was surprised at how many people in 2019 had salaries over $200K. It really spikes up and I wonder what the reason is for that