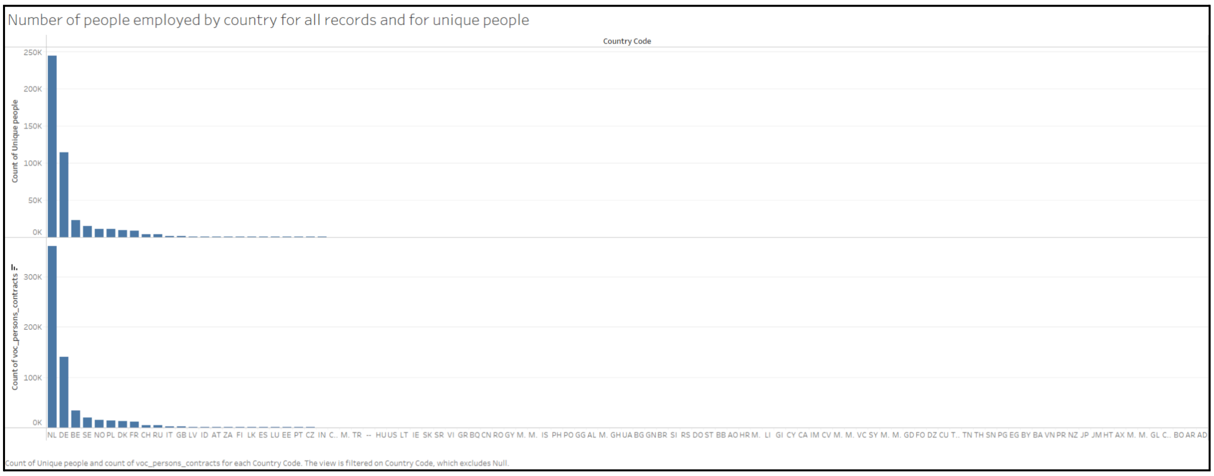
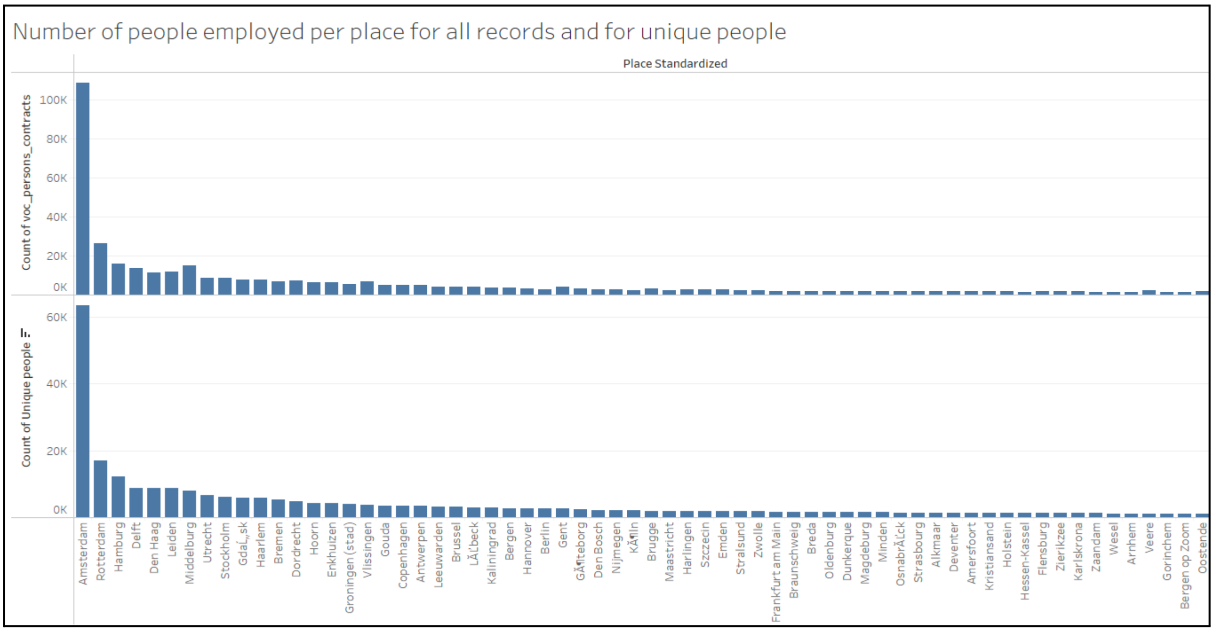
We started out by just playing around with visualisations in Tableau.



Number of people employed by country for all records and for unique people

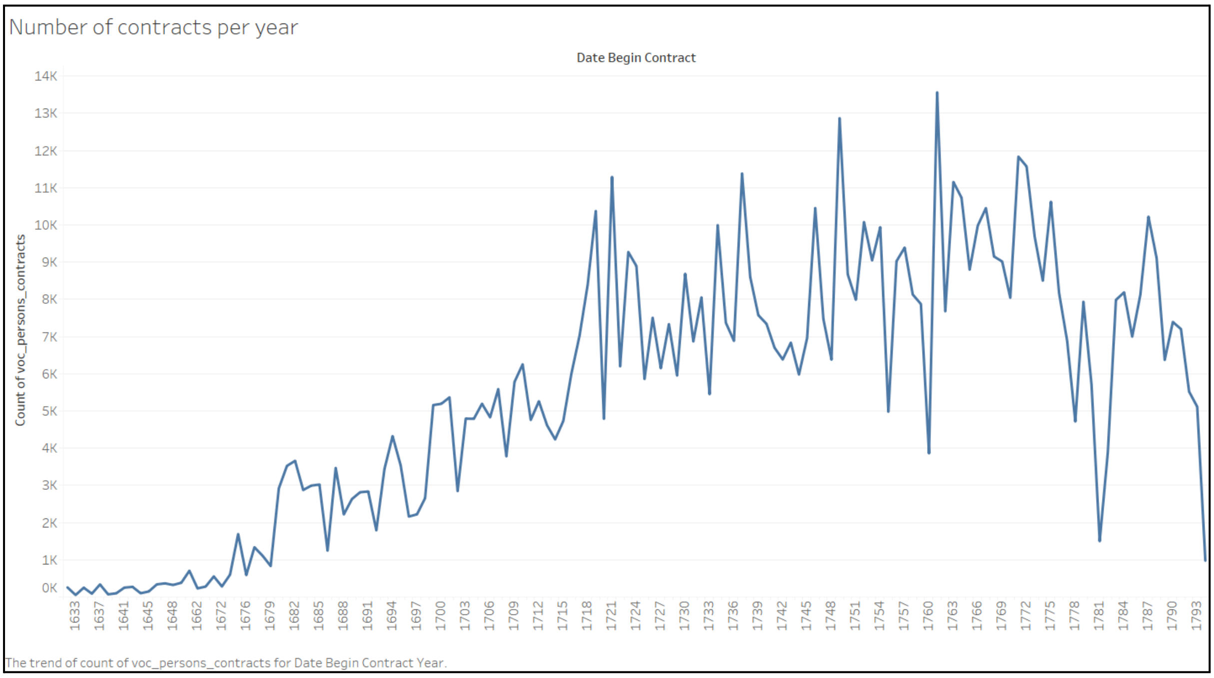
We started by seeing where the people employed by the VOC came from. As we had predicted, most of the people came from the Netherlands.
We later returned to this chart and added a distinction between all the records and unique people. The result was to note that the same distribution of people was the same country, but we found that there are fewer people.
We then distinguished unique individuals by looking at which person cluster id and taking only their first vocop id as the unique vocop id. This made sure that all their grade and salary information was only from their first contract. Although this is necessary to obtain an accurate population count, it may cause some problems in the wage data later.



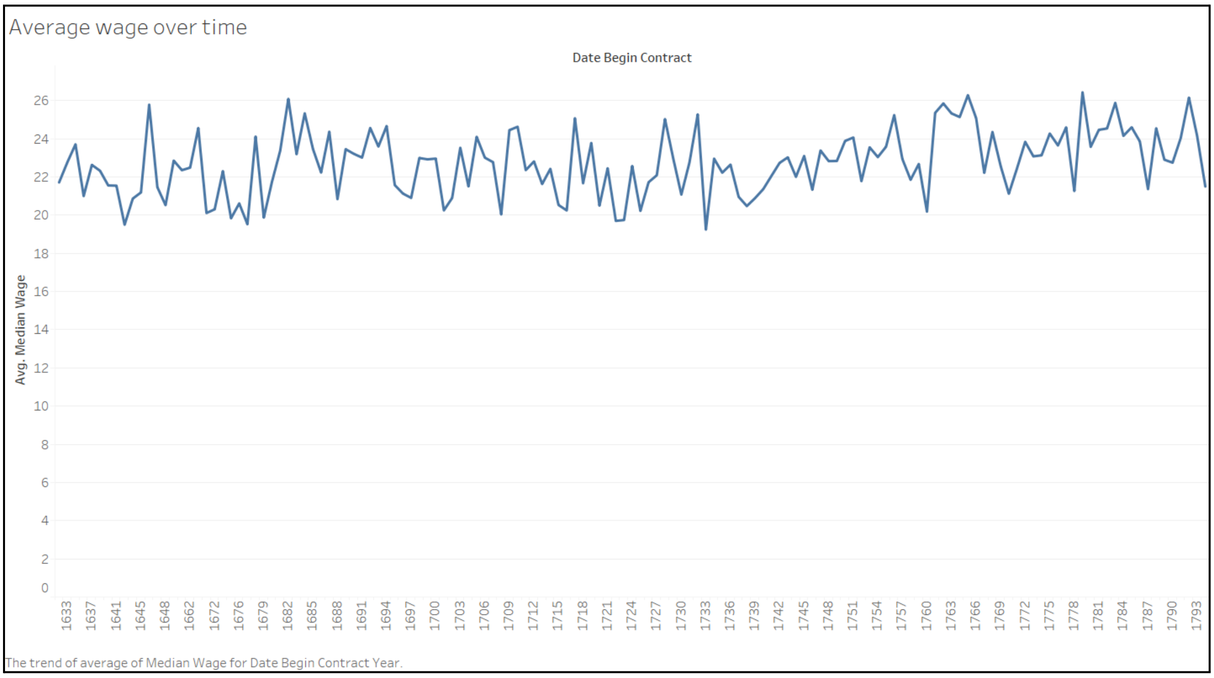
Number of people employed per place for all records and for unique people

Setting people by their origin (local edition).
As expected, the largest cities had the most VOC employees

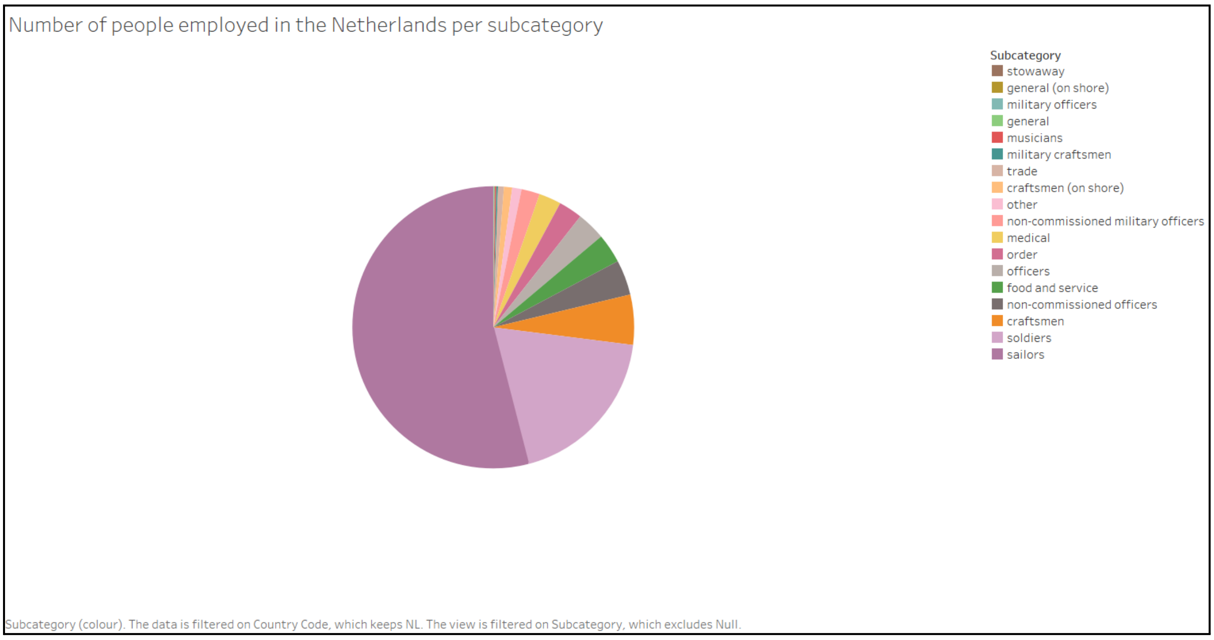
We processed the same comparison but with unique people.



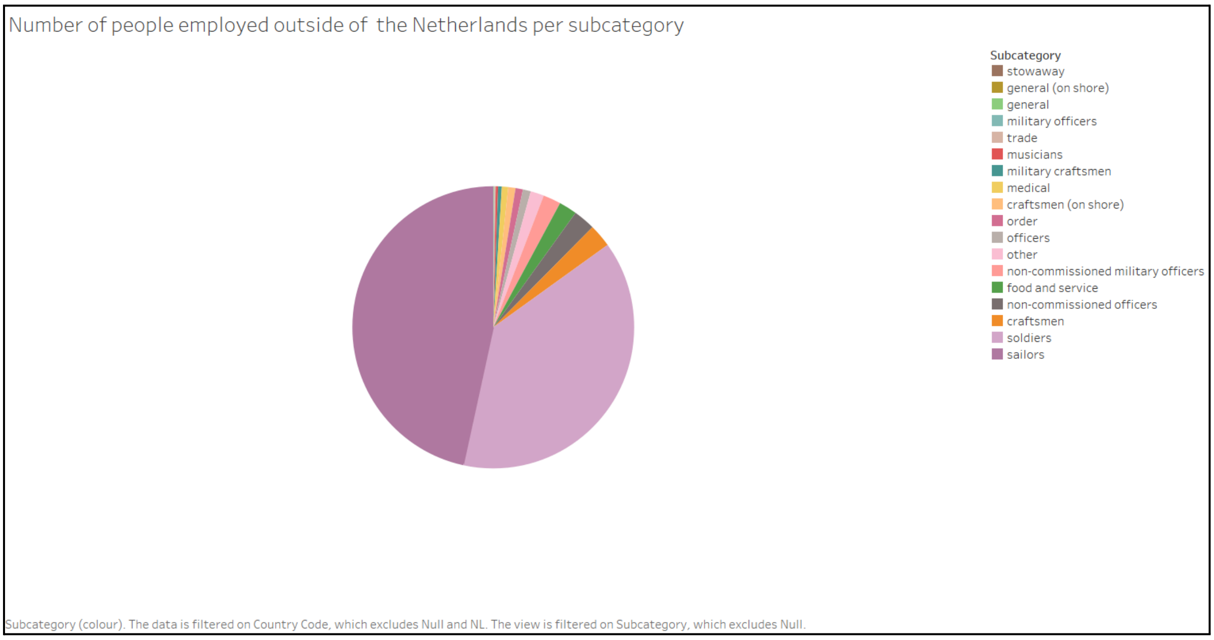
- We wanted to see how many people the VOC employed over time
- There is data missing at the first period
- Graph shows a steady increase in the number of contracts



- Average wage remained steady over time, indicating the division by categories remained roughly the same
- The limitation of this dataset is that it doesn't take into account inflation, which could be important considering this is a time period od 150 years



-

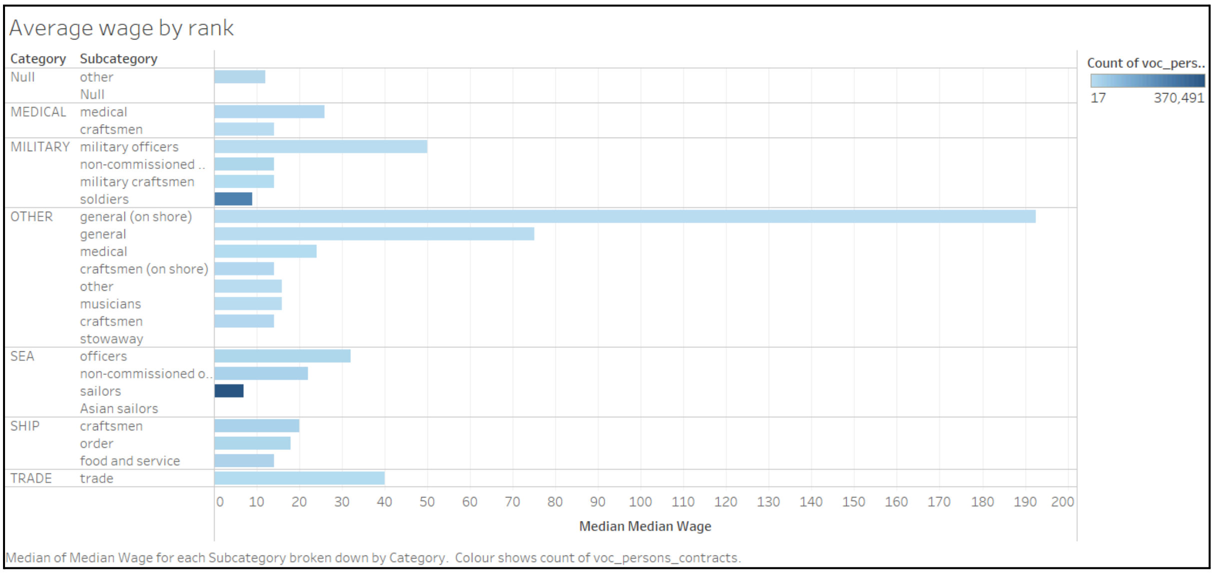


We wanted to see what the composition of the workforce was in the different countries. So we decided to compare the Netherlands with the rest of the world. However, since there are too many degrees, we decided to make a pie chart of the subcategories. Subcategories are better than categories because categories would hide too much. Too many data make a category's picture.

It can be seen that the categories of sailors and soldiers were the most numerous.

The comparison showed that the Netherlands has a lower proportion of sailors and soldiers and also that the Netherlands has a higher availability of jobs at higher levels.

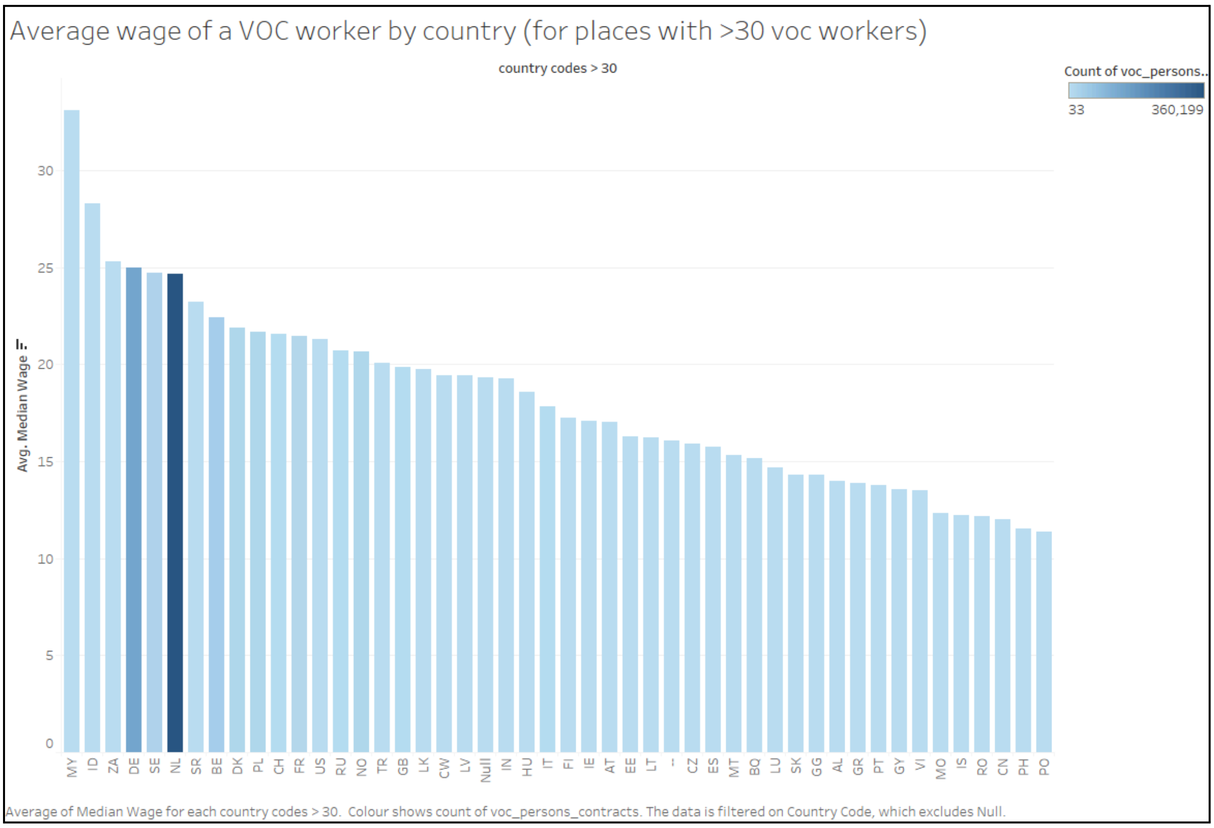
This led us to reflect on how we might consider the fact that a job was of a higher grade. We came to the conclusion that the easiest way to do this was to compare the wages they earned.



Median of Median Wage for each Subcategory broken down by Category. Colour shows count of voc_persons_contracts.

We wanted to compare the average median wage by rank

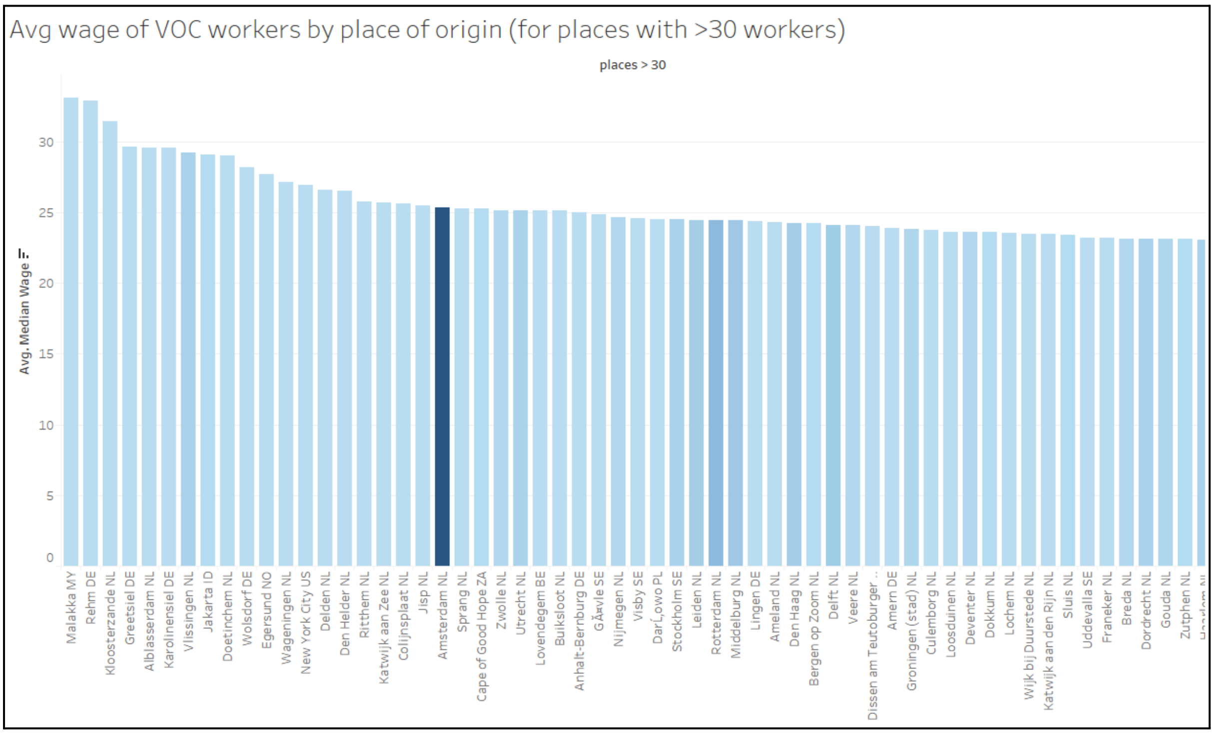
- Decision: why the average and not median? We decided to use the average median wage for the rest of the calculations. We decided it since it showcases the distribution across ranks better. The median would depend too much on the amount of people in each rank by country
- Again, w decided to compare it by subcategory
- The colour shows which subcategories employ the most people: sailors and soldiers

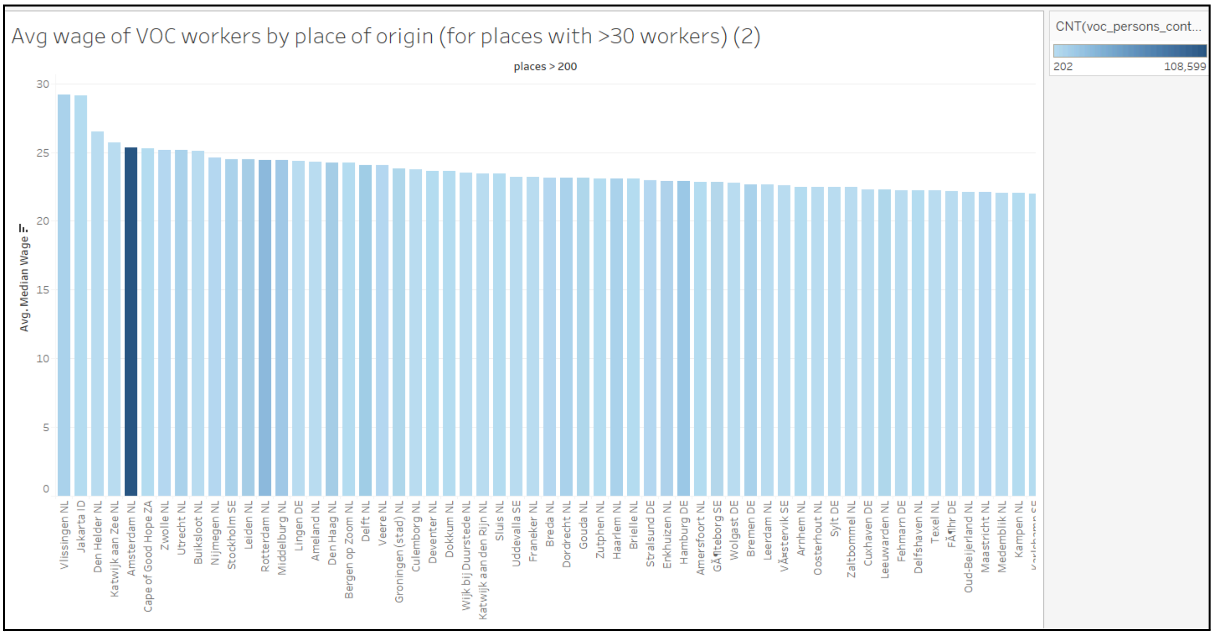


Average of Median Wage for each country codes > 30. Colour shows count of voc_persons_contracts. The data is filtered on Country Code, which excludes Null.

- average wage by country, with colours representing the number of people from the country

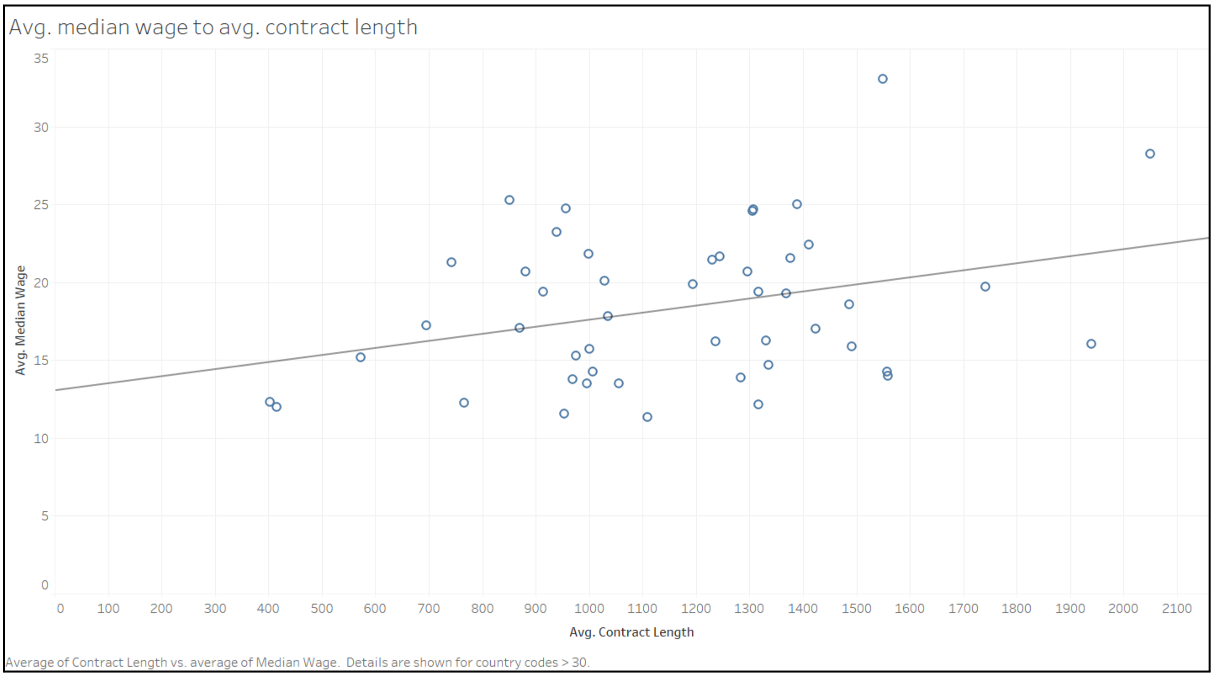
- we filtered for countries with more than >30 people employed in the VOC - why? Boris said 30 makes it statistically significant. Otherwise countries with just a few people employed in high ranking jobs skew the data.
→ Trade off: Significant results without obscuring data? We need to draw the line at a certain number of employees - the line being too low will obscure the results, i.e. countries with a few VOC employees where one happens to be higher ranking will drastically outperform (e.g. TW has 3 employees), but drawing it too low will make it so that we only focus on the biggest countries - losing a lot of insights
- The Netherlands, Germany and Belgium, being the biggest countries, still have relatively high wages compared to the rest. However, a few Asian countries still made it to the top. Why could this be? Perhaps there were more low paying jobs/enslaved people from those areas who were deemed less important to document, whereas the higher ranks were well documented.



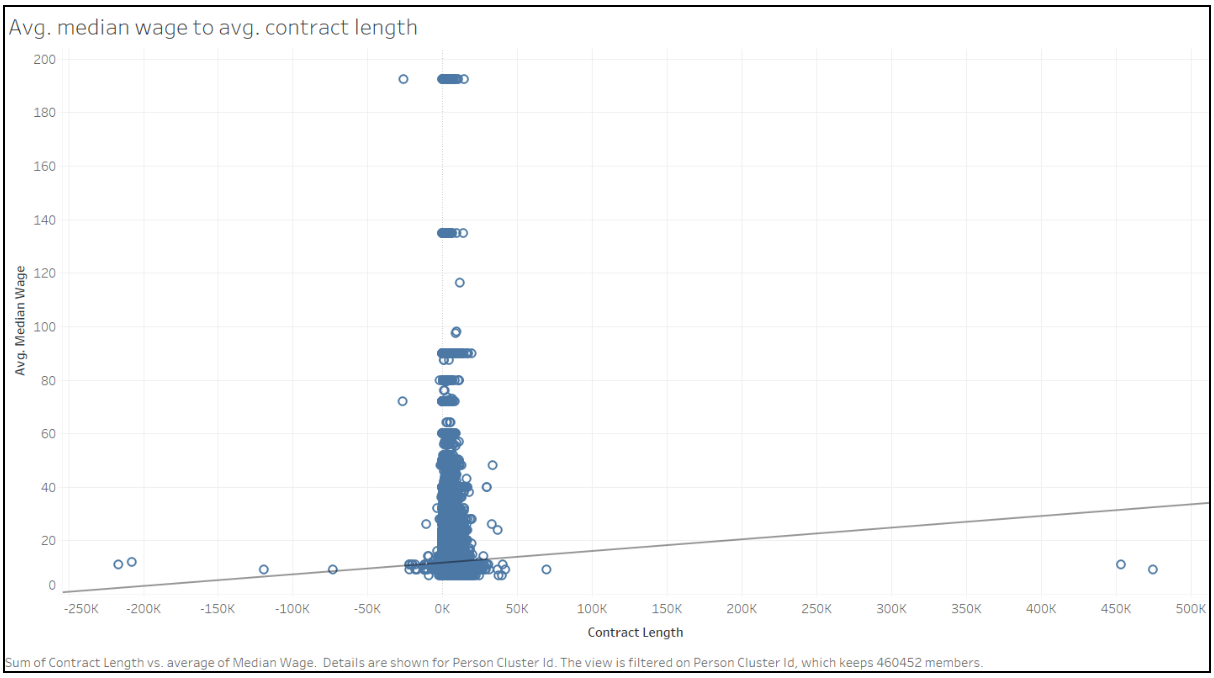


Now we do the same for places.

- Tradeoff: where do we draw the line?
- 1st graph for places >30, 2nd graph for >200 people.
- Result: The bigger cities have a higher average wage

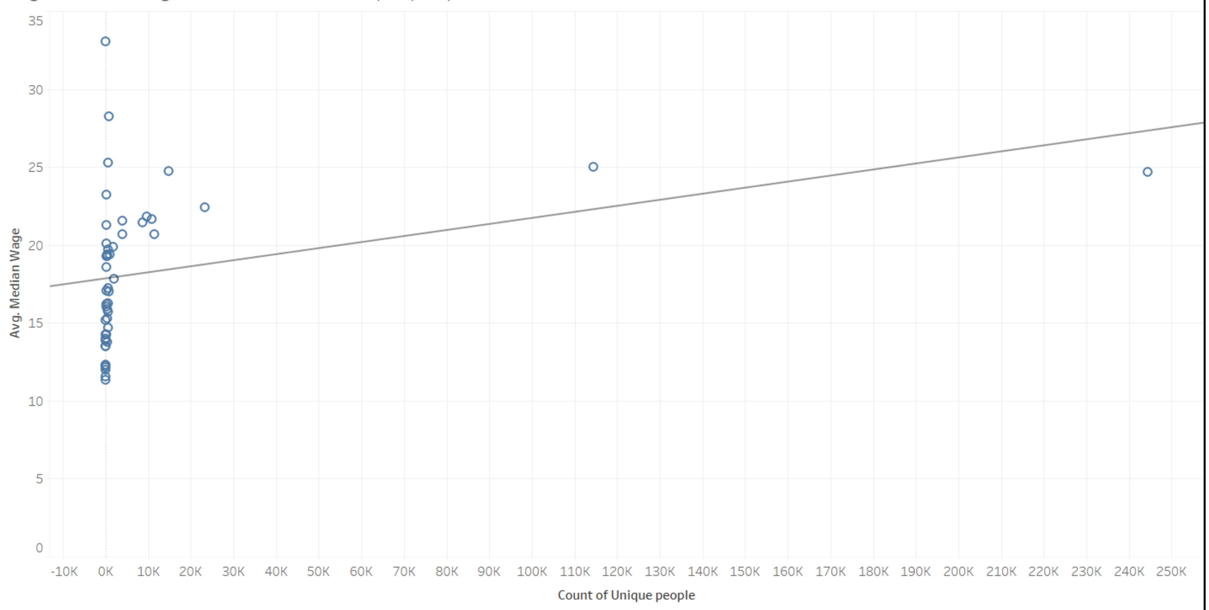


Average of Contract Length vs. average of Median Wage. Details are shown for country codes > 30.



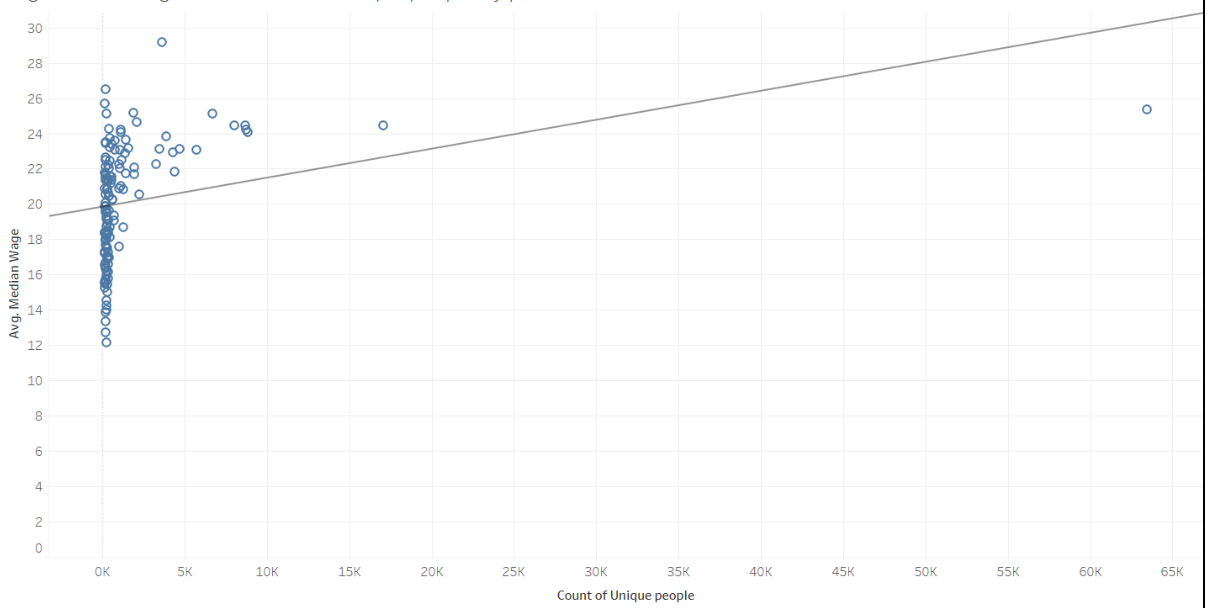
- We wanted to see whether the length of a person's contract influences their avg. median wage.
- The first graph shows the average length or contracts per country - we found no significant results, but see the average length per country was a bad measure - it has too many different contracts to make this relevant
- The second graph tries to map this for each employed individual but is completely unreadable
- Some contract length fields just randomly have minus numbers ??? a few people have -200,000 listed as their contract length? These are flaws in the design of the dataset. Some inspection yielded that there are typos in the dates of the beginning and end of contract. We could manually correct them, which would be labour intensive and could lead to errors, or simply ignore them - which would take away a lot of nuance from our dataset. In the end, our research went in another direction so we did not have to tackle this problem
- This could've been done better with some statistical data analysis, and not purely visuals.

Avg. Median wage to the count of unique people

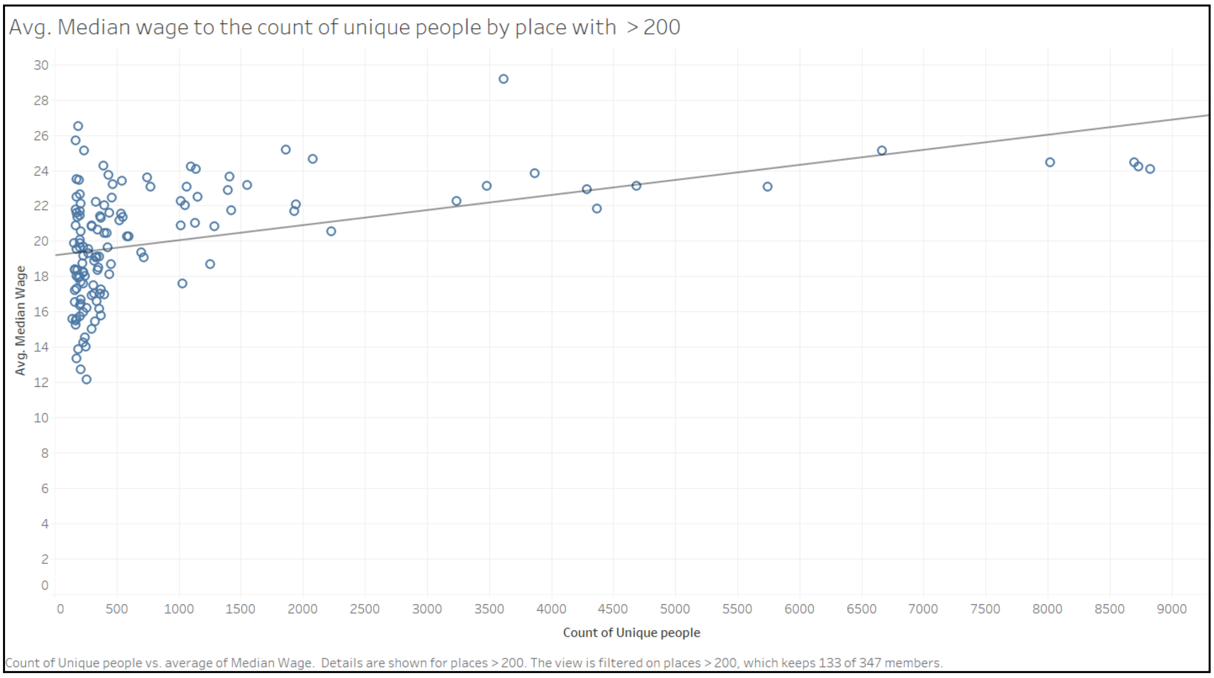


Count of Unique people vs. average of Median Wage. Details are shown for country codes > 30. The data is filtered on Country Code, which excludes Null.

Avg. Median wage to the count of unique people by place with > 200



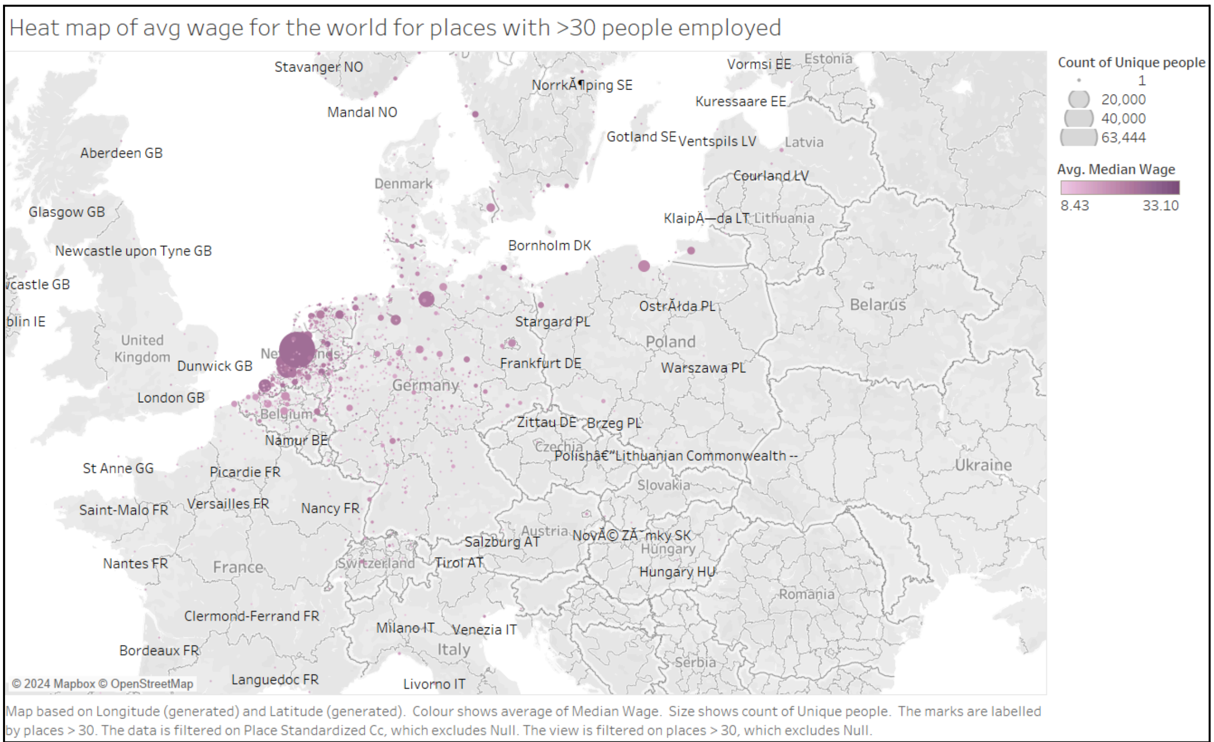
Count of Unique people vs. average of Median Wage. Details are shown for places > 200. The view is filtered on places > 200, which keeps 135 of 347 members.



Count of Unique people vs. average of Median Wage. Details are shown for places > 200. The view is filtered on places > 200, which keeps 133 of 347 members.

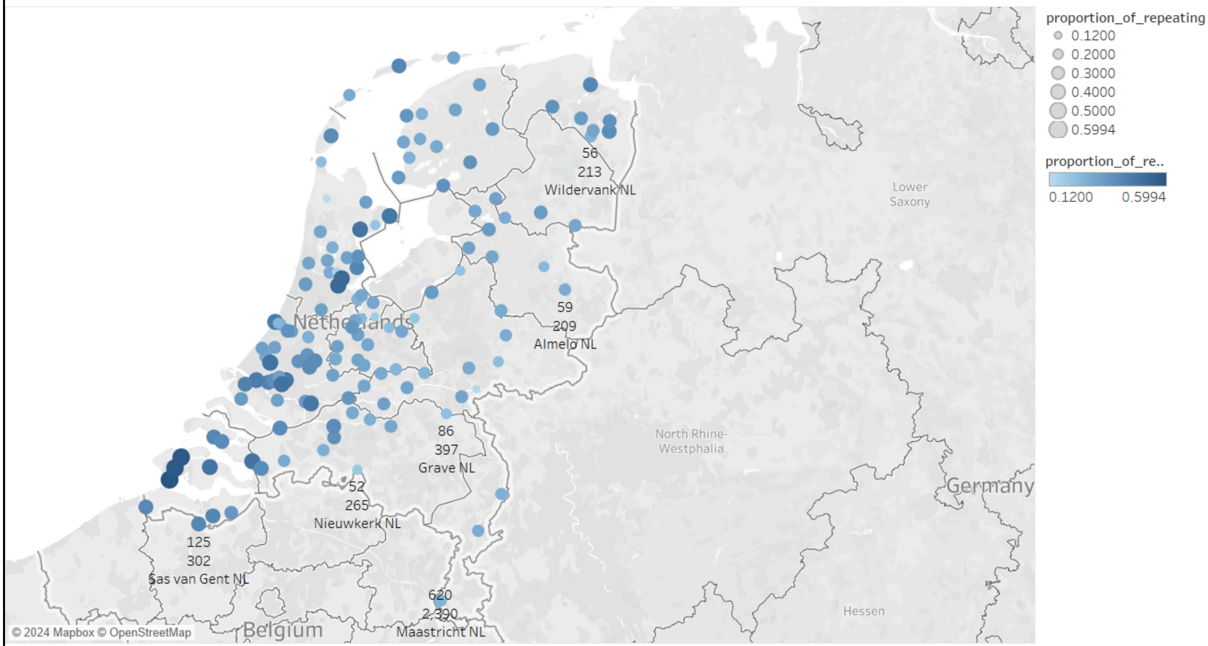
- Similarly, the avg median wage shows a slightly proportional relationship to the count of unique people from each country. The Netherlands and Germany make this graph hard to read since they are such extreme outliers. Removing the countries with a low count of employees also makes the graph a bit clearer, but it's still does not seem significant
- The first image is by country and the second by place in the Netherlands. The third is excluding Amsterdam and Rotterdam since they are such extreme outliers. Slight positive trend
- We wished to notice more patterns rather than just a positive trend.
- Maybe look for characteristics of different cities? We could make a heat map

- Heat map of proportion of repeating contracts in the NL
- Matches the bar chart Jorge made, more repeating contracts near the coast and in this one little island - we will research why that is
- We made an assumption that re-employment usually means going up in rank

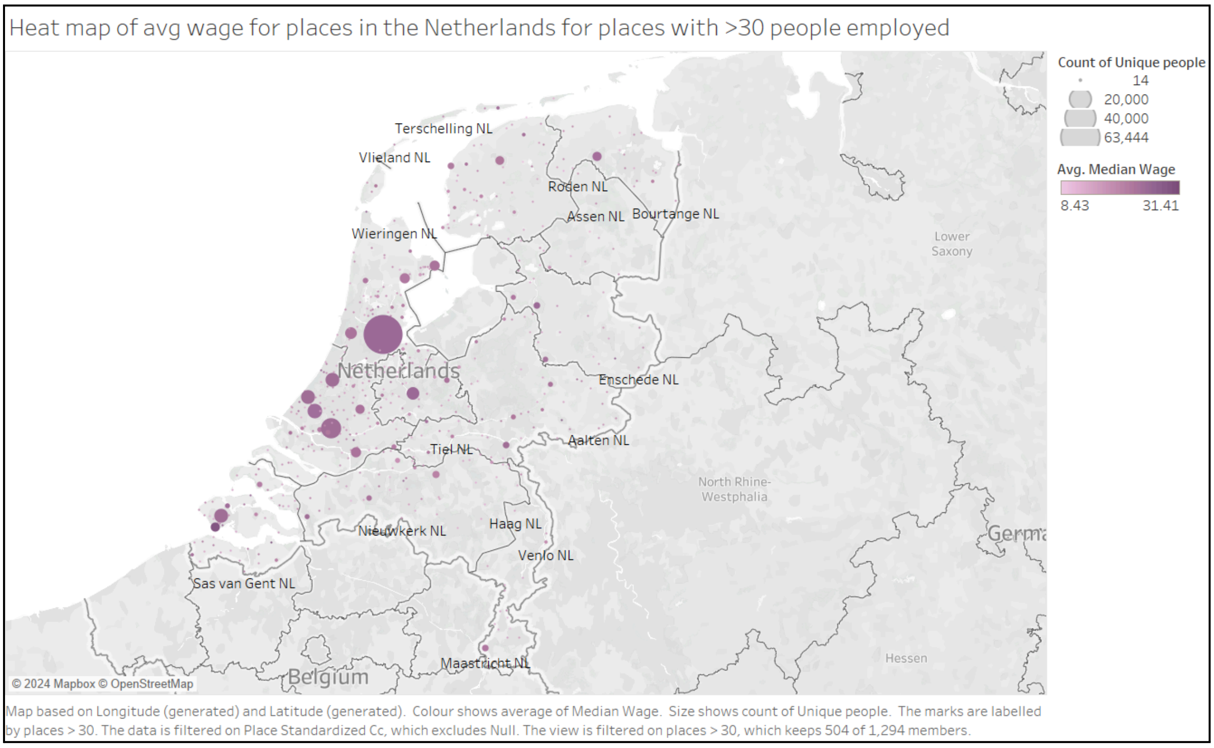


Map based on Longitude (generated) and Latitude (generated). Colour shows average of Median Wage. Size shows count of Unique people. The marks are labelled by places > 30. The data is filtered on Place Standardized Cc, which excludes Null. The view is filtered on places > 30, which excludes Null.

Heat map for proportion of repeating contracts in the Netherlands

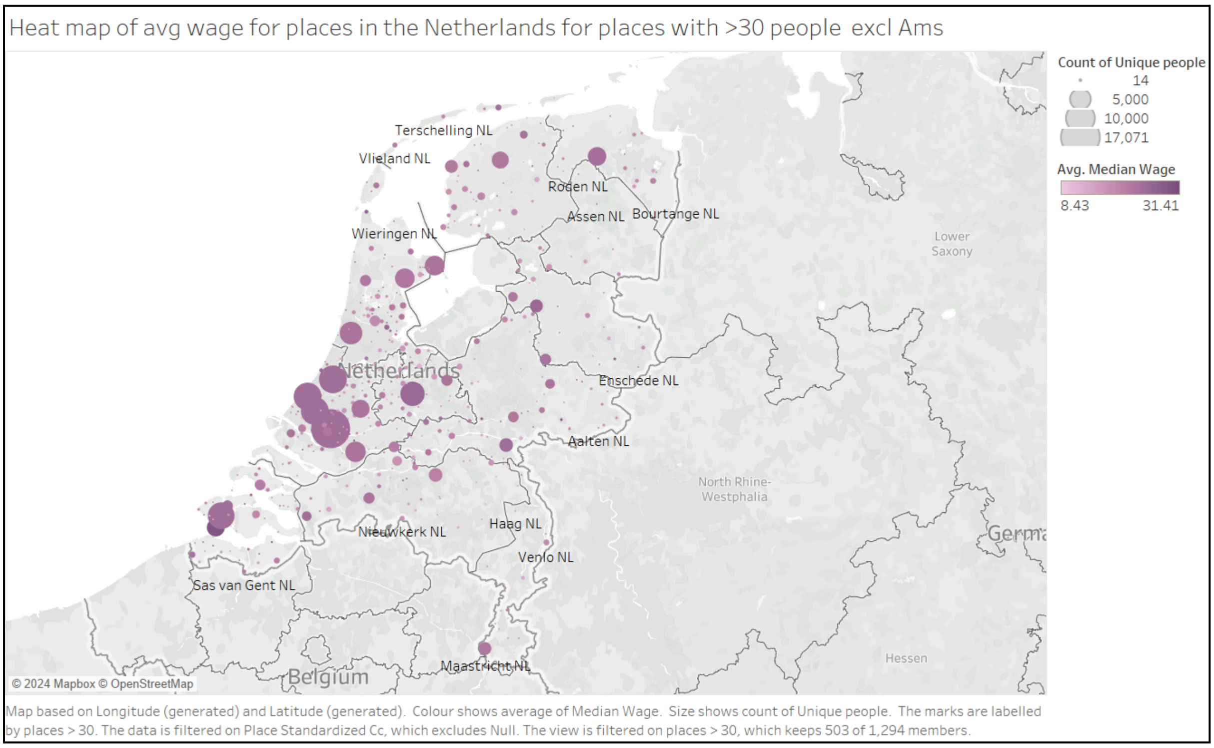


Map based on Longitude (generated) and Latitude (generated). Colour shows proportion_of_repeating. Size shows proportion_of_repeating. The marks are labelled by count of Repeating people, count of voc_persons_contracts and places > 200. The view is filtered on places > 200, which keeps 135 of 347 members.



Heat map of avg wage for places in the Netherlands for places with >30 people employed

Map based on Longitude (generated) and Latitude (generated). Colour shows average of Median Wage. Size shows count of Unique people. The marks are labelled by places > 30. The data is filtered on Place Standardized Cc, which excludes Null. The view is filtered on places > 30, which keeps 504 of 1,294 members.



Heat map of avg wage for places in the Netherlands for places with >30 people excl Ams

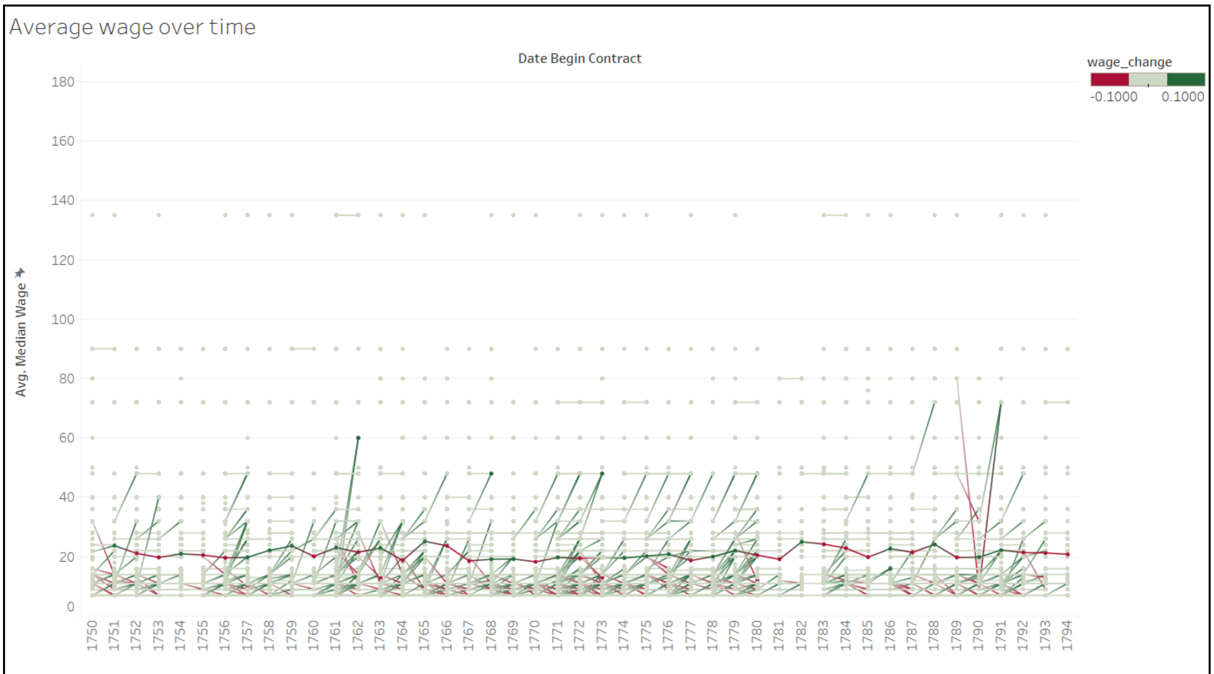
Map based on Longitude (generated) and Latitude (generated). Colour shows average of Median Wage. Size shows count of Unique people. The marks are labelled by places > 30. The data is filtered on Place Standardized Cc, which excludes Null. The view is filtered on places > 30, which keeps 503 of 1,294 members.

- Heat map of median wage (color) and number of unique people from the place for places with >30 people. We decide to take 30 to make the results statistically significant (normal distribution)
- This turned out to maybe not be the best way to represent the data since in a picture it is hard to see the color of the smaller dots, unlike in Tableau where its easy to zoom in. We could have just use color or size and left out the other dimension, but we wanted to see the correlation of the wage and number of people employed. The

benefit of this analysis is that it doesn't only show the correlation but maps it, so we can notice trends like big/small cities or land/coast. Moreover, since the range of the wages is actually rather narrow, it just made for an uninformative graphic - the sizes of the dots were too similar in size.

- We made one map for the whole world - but it was unreadable since there is so much data from the NL area that either those dots obscure the whole region or the smaller dots are unreadable.
- We then focused on just the area of the Netherlands - or at least what is considered the borders today. Again, as the biggest city, Amsterdam obscured too much. So, we made another graph excluding it to showcase different places that stand out. Again, the island stood out in both the number of people employed and the avg. median wage

- Finally, we wished to test our hypothesis that re-employment means going up in rank. This made sense, and some research seemed to back it up. We did not really know how to approach this. First, going up in rank is assessed most simply by looking at the wages, but research suggests that that might not paint the whole picture. Moreover, looking at just the increase in pay over time shows no significant trend and is not that informative since it cannot distinguish between the overall trend and trend in each re-employed individual. So, we cannot differentiate between a person who went up/down in rank, just new contracts. I did not know how to fix this and compute this on just individuals. Even so, that analysis could not take into account how many times people were employed, how likely the pay rise was, at which reemployment it happened, between which ranks… I decide to make the following:



The trend of average of Median Wage for Date Begin Contract Year. Colour shows wage_change. Details are shown for Person Cluster Id. The view is filtered on Date Begin Contract Year, which excludes 96 members.

- Each dot represents a reemployed individual. I had to limit the graph to a 40 year time period since the dataset is too large to work with in such a graph.
- The lines connect contracts of the same individual. Increases in pay are coloured green, and decreases red.

- Once again, the visual is more informative when the team can zoom in and inspect the lines more closely.
- We concluded that employees were slightly more likely to get a pay raise, especially when it is a significant amount, than a pay cut.
- The most changes happened between the lower-paying jobs, with slight raises/cuts.
- Overall, this was not the best nor the most presentable visual to base our conclusion on, but more precise methods would not take into account each person's journey.