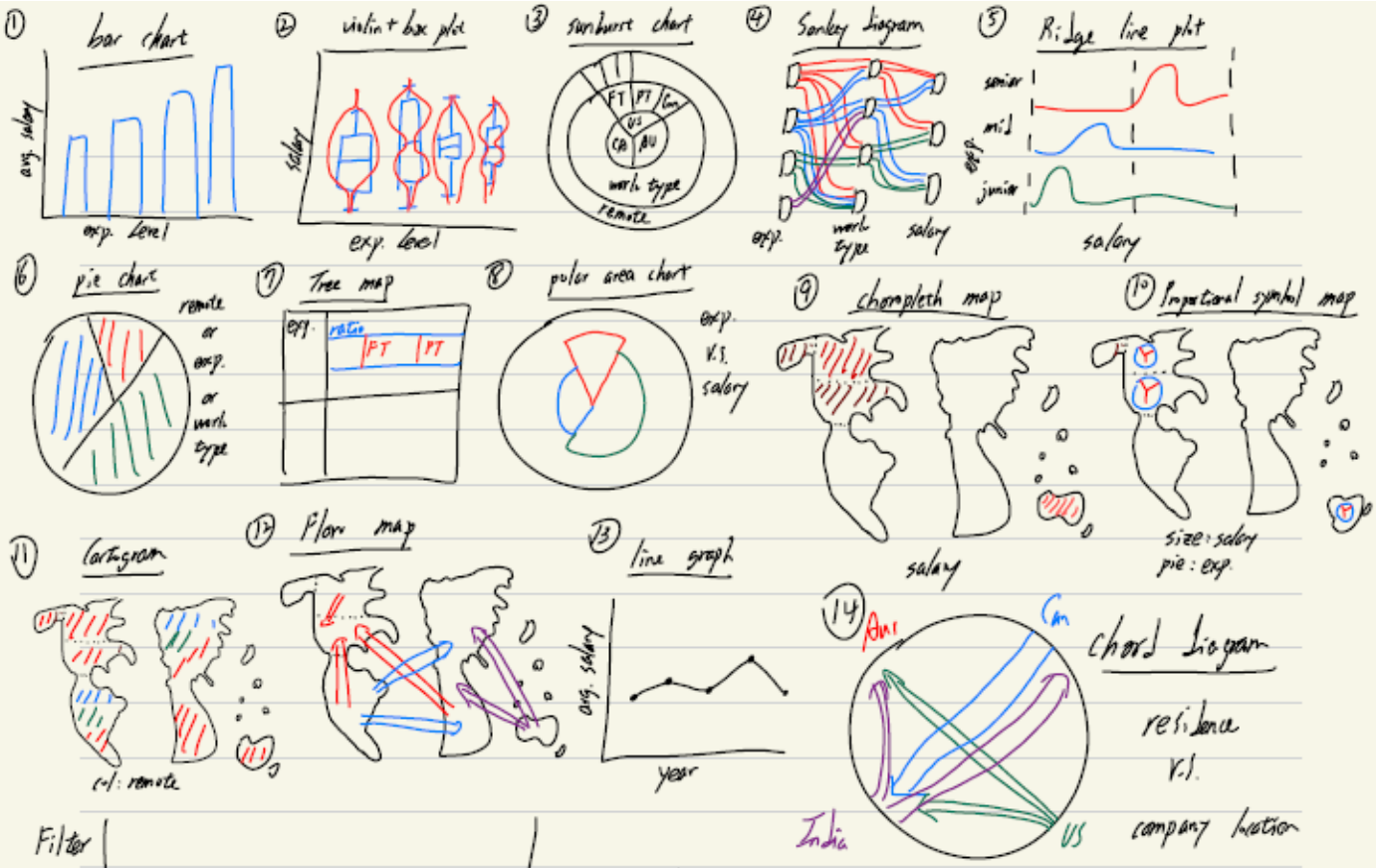
Name: Yu-Jung Ho (Malone Ho)

Title: Comprehensive Analysis of Data Science: 2020-2024

Aims: To analyse the main factors which can mostly affect salary.

Motivation: Being a student who is studying data science, I personally think it is necessary to understand any potential career opportunities before the graduation from Master of Data Science. Furthermore, it is beneficial to know average salary structures for these jobs. Hence, I am interested in knowing the relation between the amount of salary and other potential factors.



Filter

1, 6, 13 can use other complex graphs to present.

12, 14 are similar ⇒ 12 will be filtered.

3, 7 are similar ⇒ 7 will be filtered.

⇒ 2, 3, 4, 5, 8, 9, 10, 11, 14 will be kept.

Categorise

Map: 9, 10, 11

Statistical: 2, 3, 5, 8

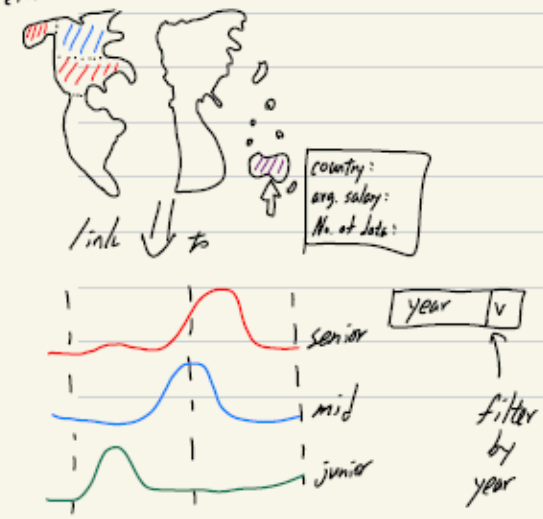
Flow or Relation: 4, 14

Title: Comprehensive Analysis of Data Science Salaries: 2020~2024
Author: Malone Ho
Sheet: 1

Combine + Refine

* Map can be interactive and linked to other statistical graphs.

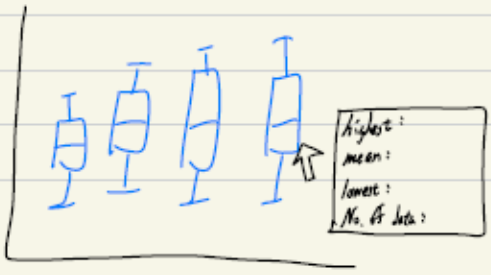
ex:



* Statistical and flow type of graph can become animation to show the changes through years.

* Statistical and flow type of graph can have filter to show the data in specific year.

* Some statistical graphs can be interactive by showing tooltip which contains detailed info.

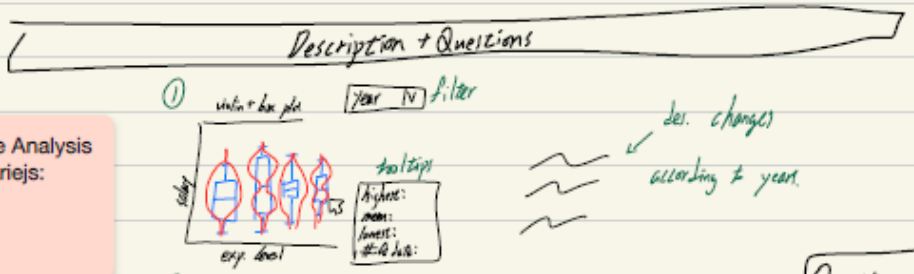


Title: Comprehensive Analysis of Data Science Salariejs: 2020~2024
Author: Malone Ho
Sheet: 2

Comprehensive Analysis of Data Science Salaries: 2020 - 2024

By Malone Ho, date

Description + Questions



Pros:
- Use various graph to present info from dataset.
- It doesn't required any prior knowledge to read this website.

Cons:
- This website might be too long, which needs readers to keep scrolling to read.

Operation:
① Tooltips show when hovering the specific boxplot.
Filter changes the focus to the year. (2020~2024)
Description changes according to the selected year.
② This diagram changes through years by animation.
③ Filter changes the presentation way, either diagram or map.
④ Tooltips show when hovering the specific country.
⑤ This graph is linked to the country that user selected.
⑥ This matrix uses motion to highlight the specific column.
⑦ Static decision tree.

Focus/Zoom:
The first description is a brief introduction and findings of the dataset. This dataset will be introduced from simple to complex. Thus, exp level will be introduced first in ①, followed by ② which explains the relationships between exp level, work type and salary.
③ explains the relationship of employees whose residences aren't same as company location.
④ presents the avg salary in different countries. Also shows the distribution of salary separated by exp level in ⑤ which is linked to ④.
⑥, ⑦ are the results of comprehensive analysis among all variables. ⑥ shows the correlation between each variables and ⑦ is the decision tree to identify the intervals of salary.

Comprehensive Analysis of Data Science Salaries: 2020 - 2024

By Malone H., date

Title: Comprehensive Analysis of Data Science Salaries: 2020~2024
Author: Malone Ho
Sheet: 3

Operations:

Same as sheet 2.

+④ changes through years by ani.

+⑤ has filter to chose how many 6 countries are shown in the map.

Focus/Zoom:

Same as sheet 2.

+ ④ presents the detailed distribution under different attributes.

Pros:

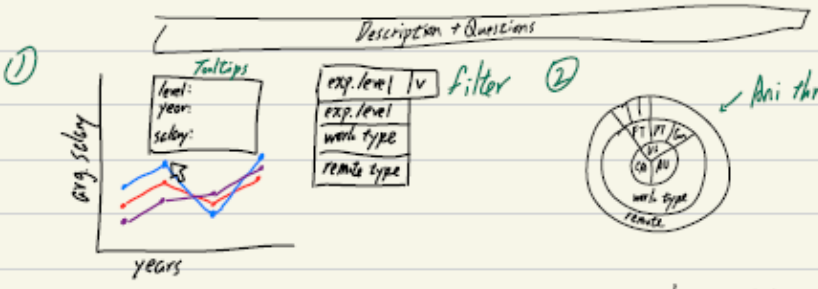
- Symmetric website
- Shorter website than sheet 2

Cons:

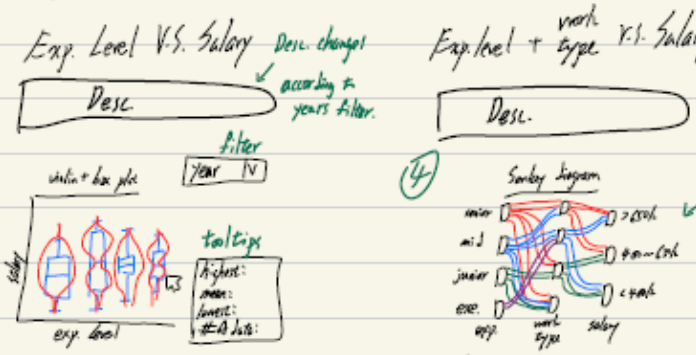
- ④ might be too complicated to understand since it includes too many information.
- This website might be too crowded.

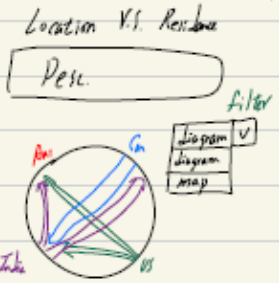
Comprehensive Analysis of Data Science Salaries: 2020 - 2024

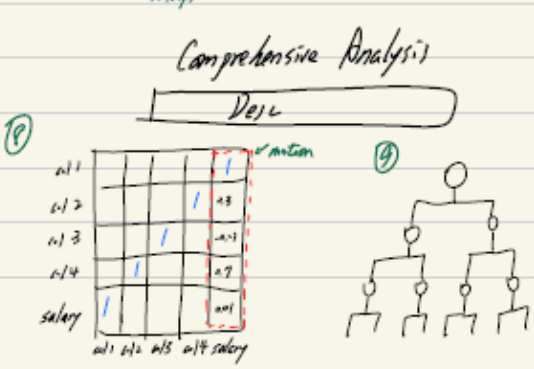
By Malone H., date



Title: Comprehensive Analysis of Data Science Salaries: 2020~2024
Author: Malone Ho
Sheet: 4







Operation:

① Tooltips show when hovering on specific point and it shows the detailed info of the point.

Focus/Zoom:

① allows readers start understanding data from simple to complex.

Pros:

- From simple to comprehensive
- Bigger space for map section ⑤.
- Symmetric

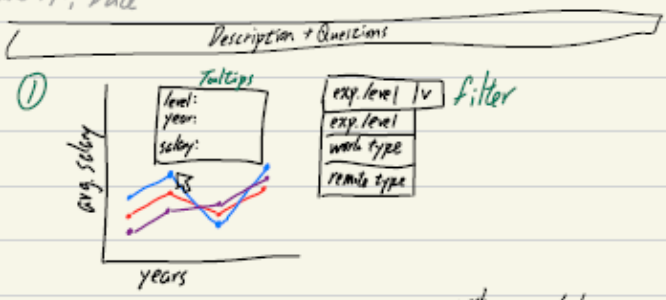
Cons:

- ② might be too complex to understand since it includes too many info.

Comprehensive Analysis of Data Science Salaries: 2020 - 2024

By Malone Ho, Date

Title: Comprehensive Analysis of Data Science Salaries: 2020~2024
Author: Malone Ho
Sheet: 5



Detail:
The whole website will be constructed by 8 mvs.
① use eventReactive() in shiny to chang graph based on the filter.
② use gganimate to present animation
③ use networkD3 and gganimate to implement
④ use circlize library to implement chord diagram and leaflet to present interactive map
⑤ use eventReactive() to change data shown in map, leaflet lib to present reactive map.
⑥ ggridges to create the graph.
⑦ use corr() to create correlation matrix
⑧ use rpart() to create decision tree

Exp. Level V.S. Salary

Desc.

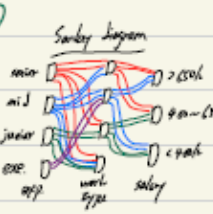
②

Ani. trough years

Exp level + work type V.S. Salary

Desc.

③



Ani. through years

Location V.S. Residence

Desc.

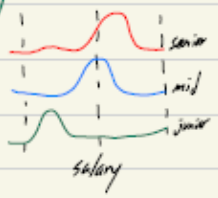
④

Location V.S. Salary

Desc.

⑤

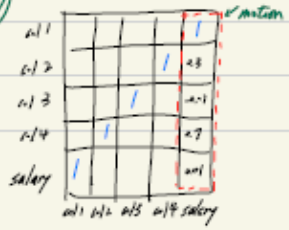
⑥



Comprehensive Analysis

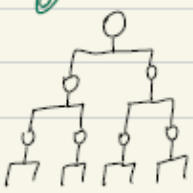
desc.

⑦



desc.

⑧



Operation:
① Use filter to include different attributes in the line graph.
②, ③ No operation required.
④ Use filter to chose either map or chord diagram shown in the section.
⑤ Use filter to select how many countries are shown in the map.
⑥ is linked to ⑤. Hence, the info in ⑥ is depending on the country that selected in ⑤.
⑦, ⑧ No operation required. ⑦ includes motion to highlight the salary column.

Focus/Zoom:
①, ② let readers can understand the data generally. Then these two graphs are followed by detailed and comprehensive graphs and maps.
③ provides the relationships between exp level, work type and salary.
④ reveals the flows from residence to company location.
⑤ shows the avg. salary in different countries and
⑥ present the detailed salary distribution in diff. exp. level.
⑦ shows the correlation coefficients between each pair of variables.
⑧ shows the most significant variables that affect salary.