

Lanzhou University

INFO 250: Information Visualization

Project 2

Group 23

Haiyi Wang: 320200931281

Siyang Xie: 320200931361

Zijun Yu: 320200931500

Longhao Li: 320200931051

How to build a simple, understandable visualization? An improvement of Nightingale rose diagram

1. Abstract

Data visualization perfectly combines technology and art to help us extract knowledges from information and then harvests values from knowledges. In this report we found a visualization that uses a **Nightingale rose chart** to represent certain aspects of sales in China.

Firstly, we introduce and analyze the importance of the original visualization, briefly enumerate the disadvantages of the original visualization and the reasons for our improvement. **Secondly**, we introduce in detail the use of the original visualization, advantages and disadvantages as well as the connection between original visualization with us. We also use **cognitive theory** to analysis some aspects of the visualization. **Then**, we explain in detail how we improved the original visualization with **bar plots** represented by **small multiples**. **Finally**, we summarize the whole report and give some suggestions.

2. Introduction

Sometimes, you shouldn't trust your eyes too much. Like now, you might think I'm about to tell a strange story, but it's all about visualization. Data visualization is very important. It can help people quickly understand the importance of data and the meanings behind the data by graphing the data.

But more often than not, **bad visualizations can intentionally or unintentionally mislead the reader** without the reader realizing it, because of the trust people have in the preconditions of visualizations. For this reason, our team take a typically not so good visualization and improve it to make it more reliable to the data and better to convey more information beyond the data itself. We hope our work will enlighten other visualization writers and make readers more aware of these visualization pitfalls during our modification progress.

The [original visualization](#) illustrates data on the sales volume changes of different aspects in China from 2012 to 2021. The original figure uses Nightingale rose chart to show the comparison of sales volume of different elements in different years and the changes of each component. The **reason** why we choose it is that it is an illustration of China Daily which is a large news platform. It describes the sales changes of different aspects in China and has a large reading group and volume. Many shortcomings of the original image will cause great misdirection to the readers, so the improvement of the original image is very important and necessary.

Looking at the original image, we can find that it is presented in a complex way, which takes the reader a lot of time to figure out what it is meant to express and show. In addition, it represents data relationship not obviously, contains deception components, shows data not clearly and so on. That's why we wanted to make it simple, clear, and prioritized.

3. The Analysis of the Original Visualization

The original visualization(Figure 1) chooses a **stacked Polar area diagram**, or **Nightingale rose diagram**, to organize all the data together. The **visual variables** include “total retail sales of consumer goods”, “sales of commercial housing”, “sales of residential housing”, “sales of China's top 100 chain stores”, “sales of China's top 100 fast_moving consumer goods”, “Total retail sales of alcoholic products from enterprises above designated size in China” and “Total retail sales of snack foods in China”. Specifically, sales of different elements are stacked in increasing order of data size from the inner ring to the outer ring, with the polar coordinates increasing clockwise from 2012 to 2021.

Here is our **replication** of the original visualization (using Matplotlib in Python):



Figure 1. Replication

The visualization is only as helpful as it is understandable. We have to admit that this visualization is much easier to understand than the data alone. Just looking at the data makes it more confusing, while the visualized graph is clear.

From the aspect of **cognitive theory**, the visualization has its **effectiveness**:

As for **ermane cognitive load**, the connection between the audience and visualization is intentional. The audience click the news on China Daily to see it. Because of a lot of information is presented, the audience will look slow at it. The visualization only shows the sales volume changes of different aspects in China from 2012 to 2021. So even without some professional knowledge, the audience can understand what is showing. As for confidence, the novel form of the Nightingale rose diagram will make the audience feel comfortable.

From the aspect of **cognitive theory**, the visualization also has its **ineffectiveness**:

As for **extraneous cognitive load**, the visualization uses a very rare chart type, so it takes the audience a lot of time to the chart format. Due to the area amplifies the actual ratio between the data, it is approximate but not accurate. The composition of it is really complex and the information can be hardly conveyed. Besides it also uses a legend, which requires the reader to load the working memory and not focus the visualization as a whole.

In detail, this visualization also has these following **drawbacks**:

- **Misdirection**: We can see that the 4 groups of indicators located in the inner circle of Nightingale rose chart show almost no fluctuation during 2012-2021. However, these indicators changed significantly, especially Retail sales of consumer goods. Its value has nearly doubled in the past 10 years, but significant change is hidden.
- **Lack of Comparability**. We can observe that in 2014, the indicator "Total retail sales of aluminum products from enterprises above designed size in China" and "Sales of China's top 100 fast moving consumer goods" decreased compared with the previous year, but increased in the following year. What's so frustrating is that we cannot judge the size of these two indicators in 2013 and 2015 because the human eye is not sensitive to the area.

In our opinion, using Polar area diagram to organize these data is like recklessness after drinking(just a metaphor), which is probably an unwise behavior. We take the form of using **bar charts** in **small multiples** with **different sizes** of the subgraph to improve the visualization.

4. The Improvement of the Original Visualization

As a result, we make the following improvements:

1. From the **whole visualization level**, we changed the expression form of visualization. We use bar chart instead of Nightingale rose diagram. Since the relationship between radius and area is square, the Nightingale rose map visually exaggerates the scale of the data. This will mislead to the reader to some extent. By using bar charts, it can not only present the sales volume in different years but also reflect the data relationship between different elements to avoid cheating.
2. From **each element display**, we use small multiples to display each element. Several elements near the center of the circle in the original figure, such as "Total retail sales of snack foods in China", are hardly visible in the original figure due to their small radius, not to mention reflecting the changing trend of each ingredient. Therefore, we use small multiples to display each element, which can not only better reflect the changes of each element but also allow readers to analyze a certain element.
3. From the **focus and sub-focus of the visualization**, we adjust the size of different small multiples to indicate the different importance of different elements in the visualization. Although the area of different elements in the original picture reflects the importance of different elements, it is difficult for readers to observe elements close to the center of the circle. By changing the size of different small multiples, we emphasize the importance of different elements. Besides, it is simple and beautiful and enables readers to see the data of different elements more clearly.

And here is our improvement of the original visualization:

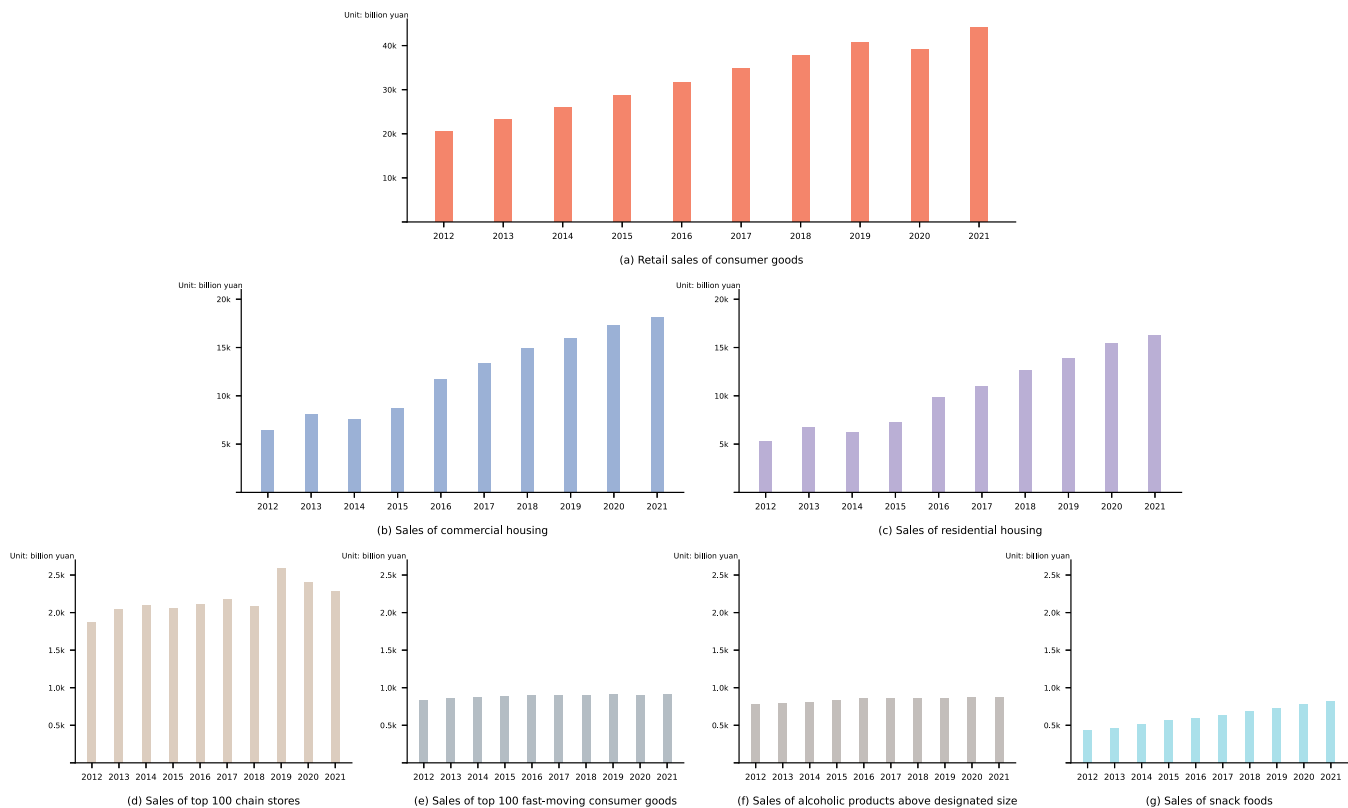


Figure 2. The Improvement

What's more, these changes could also be applied to other generic graphs:

- When the visualization uses a too unnormal graph which takes too much time for readers to understand it, the visualization can be just replaced by another normal forms.
- When the data-ink is too low, the visualization could get a higher data-ink by removing useless borders and avoiding elements that interfere with vision and so on.
- Keep real and avoid lie factor.
- When the visualization contains too many factors, it can be divided into small multiples to make is clear.

5. Advantages of the Improvement

Our visualization improves the shortcomings of the original one and has many advantages:

- Simple and clear. By using bar charts to illustrate elements of the data, our visualization looks much simpler and clearer than the original one. Readers can easily understand what we're trying to show.
- High **data-ink ratio**. In the process of visualization, we try to avoid non-data-ink and redundant data-ink. Our visualizations support the information we want to present as much as possible. We keep our visualizations at high data-ink ratio by removing useless borders and avoiding elements that interfere with vision.
- Be authentic. In order to make our visualization more objective and real, avoid to misleading readers, we correctly present the real ratio between the data and modify the **lie factor** and ambiguity in the original picture.
- By using small multiples, we can clearly observe the sales and changes of each element. For different sizes of the subgraph, we can know its importance and proportion. At the same time, we can compare the data of each element in the same period.
- Using **vector** images to improve the quality of our visualizations.

6. Conclusion

Visualization is necessary because it helps the readers understand the data and makes the data more vivid and meaningful. But visualizations can't always be trusted, due to (the limitations of the diagram itself, the author's deliberate direction, etc.), the visualizations can be misleading. As for the producers, they should be more careful in drawing and choosing the

correct visualization type to present the real data as much as possible, so as to avoid the formation of misunderstanding and misleading others. As readers, they should have their own thinking while observing and analyzing visualizations. Of course, we encourage the readers to do what we do, when you find problematic visualizations, improve them and provide more quality visualizations to the community!

We have published our work to [our Github](#). Welcome to our Github and get the sharings!