INFO 250 - Information Visualization

—— Project 2 ——

Team 3
2025.12.19

Name	Student ID	Email	Role
Weixin Kong	320230941990	kongwx2023@lzu.edu.cn	leader
Yinuo Sun	320230942370	sunyn2023@lzu.edu.cn	
Shiqi Wang	320230942440	wshiqi2023@lzu.edu.cn	
Junchen Zhou	320230942790	zhoujch2023@lzu.edu.cn	

Carbon Emission Drivers Visualization Optimization in the Yellow River Basin (2005–2020)

Content

Abstract

This study focuses on the visualization optimization of stacked bar charts depicting the six major driving factors of carbon emissions in the Yellow River Basin from 2005 to 2020, as published in Scientific Reports. The core motivation is to address the contradiction between high data value and low expressive efficiency—while the key data on basin carbon emissions carried by the chart is of great significance for policy-making, enterprise transformation, and public awareness, its information transmission is limited by design flaws. The research first implements basic optimization measures such as original chart reproduction, legend migration, and bar adjustment, followed by advanced schemes including scatter plot reconstruction and interactive function enhancement. The results significantly improve the information transmission efficiency of the chart. The core insight is that high-quality visualization must adhere to data authenticity and follow cognitive laws, providing a replicable paradigm for the visualization of similar complex data.

Keywords：Carbon Emissions，Visualization Improvement，Cognitive load

---

1. Introduction

Against the backdrop of deepening global climate governance and China’s "**dual carbon**" goals, the Yellow River Basin, as an important **ecological barrier** and **economic growth pole**, its carbon emission driving data holds significant practical guiding value for formulating **emission reduction policies**, promoting **green transformation of enterprises**, and fostering **public environmental awareness**.

Information visualization is a key link in the efficient transmission of complex environmental data, and its quality directly affects the realization of **data value**. Currently, the visualization of high-value environmental data often suffers from **unreasonable design**, leading to low information transmission efficiency and difficulty in understanding for non-professional audiences.

This study takes the "stacked bar chart of six driving factors of carbon emissions in the Yellow River Basin (2005–2020)" published in *Scientific Reports* (a sub-journal of *Nature*, link: https://www.nature.com/articles/s41598-023-40998-6) as the research object. The chart has flaws such as separated legends from data, which affect reading efficiency. Following the research framework of "background combing and cognitive analysis—original chart reproduction—phased optimization—value summary", the optimization scheme focuses on reducing **cognitive load**, covering basic improvement measures such as legend integration, as well as advanced optimization strategies including chart reconstruction and interactive design. The aim is to enhance the **communication effect** and **practical application value** of environmental data visualization.

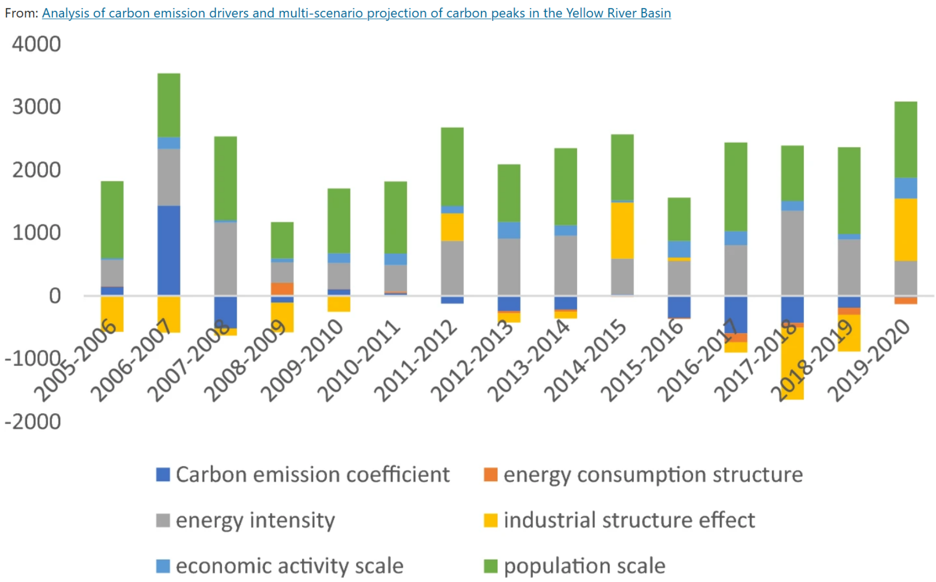

Figure 1: The original chart

2. Research Background

This study focuses on the **driving mechanism** of carbon emissions in the Yellow River Basin from 2005 to 2020, corely presenting the dynamic synergistic impacts of six factors—**population size**, **energy intensity**, **economic activity level**, **carbon emission coefficient**, **industrial structure effect**, and **energy consumption structure**—on basin carbon emissions.

The selection of this chart for optimization is based on two reasons: first, **high data value**—the Yellow River Basin serves as China's ecological security barrier and economic growth engine, and its carbon emission driving data directly supports the formulation of regional emission reduction policies, the design of enterprise transformation paths, and the cultivation of public low-carbon concepts, with prominent practical significance; second, **strong optimization necessity**—although the original chart carries high-value information, it has inefficient links such as remote legends, mixed positive and negative values, lack of units, and difficulty in reading precise values. Non-professional audiences cannot extract key conclusions in a short time, making it a typical optimization sample due to the "imbalance between value and efficiency".

Interpreting the chart requires following a four-step logic: "color-legend correspondence, position judgment for direction, length evaluation for intensity, and integration of superimposed effects". Its **visual variables** are defined as: horizontal axis for time, vertical axis for action direction, bar length for intensity, and hue for distinguishing factors. The core information can be decomposed into:

- **Positive driving forces**: Population size is the primary positive force, energy intensity maintains a high proportion, and the contribution of economic activity scale expands year by year;
- **Negative inhibiting forces**: The carbon emission coefficient continuously exerts an inhibitory effect, and the industrial structure effect has significantly strengthened its inhibition since 2015;
- **Structural transformation**: The energy consumption structure showed a positive driving effect during 2005–2009 and shifted to a negative driving effect after 2015.

From a research perspective, the aforementioned pain points are universal—visualization of multi-factor dynamic environmental data often faces problems such as remote legends, insufficient visual correlation, and confusion between positive and negative values. Targeted optimization of this chart can not only release the policy and public communication potential of carbon emission data in the Yellow River Basin but also provide a **replicable and promotable paradigm** for the visualization improvement of similar achievements, aligning with the core goal of "making high-value data accessible to more people".

3. Research Process of Visualization for Driving Factors of Carbon Emissions in the Yellow River Basin

This study adopts a systematic approach to address the core issues in the visualization of carbon emission driving factors in the Yellow River Basin, progressing gradually from theoretical analysis to practical optimization. The research process is divided into five key steps, ensuring each link is based on in-depth analysis and rigorous verification.

3.1 Step 1: Problem Analysis

This stage comprehensively diagnoses the information transmission efficiency of the original chart from two dimensions—**cognitive theory** and **practical application scenarios**—to identify key flaws.

3.1.1 Cognitive Effectiveness Analysis

From the perspective of cognitive theory, the original chart exhibits **low efficiency** in information transmission, mainly due to the following problems:

1. **Cognitive overload**: The interpretation process requires completing multiple tasks simultaneously, such as "color-legend correspondence, position judgment, and scale reading", which exceeds the working memory capacity of "7±2", easily leading to information confusion;
2. **Violation of the proximity principle**: Legends are located at the bottom of the chart, far from the data area. Readers need to frequently switch their gaze for comparison, increasing **visual search costs**;
3. **Lack of Gestalt proximity**: The width of bars is smaller than the spacing, resulting in weak **visual correlation** between factor modules in the same year, making it difficult to form an overall understanding of "synergistic impacts";
4. **Incomplete data information**: The vertical axis only marks numerical values without indicating **measurement units** (10,000 tons of CO_2), making it impossible to accurately quantify **impact intensity**; the stacked chart only allows a rough understanding of the impact of each factor, with no access to **precise information**.

3.1.2 Specific Problem Sorting

Based on practical application scenarios and cognitive analysis, the original chart has the following prominent problems in information transmission:

1. **Separation of legends from the data area**: Legends are located at the bottom, far from the bar data above. Readers need to repeatedly compare up and down, leading to **frequent gaze switching** and easy confusion of factors;
2. **Insufficient visual correlation**: Narrow bars and large spacing result in scattered distribution of factor modules in the same year, making it difficult to intuitively reflect the core logic of "multiple factors jointly acting on carbon emissions";
3. **Incomplete data labeling**: The vertical axis only marks numerical scales without unit and y-axis labels. Readers can only roughly judge the impact intensity and cannot accurately know the **specific impact** of each factor on carbon emissions;

4. **Unclear theme expression**: The chart is only labeled "Figure 1", and the core theme needs to be confirmed through the explanatory text at the bottom. Readers cannot quickly obtain the core information of "Analysis of Driving Factors of Carbon Emissions in the Yellow River Basin";
5. **Mixed stacking of positive and negative factors**: Positive and negative driving factors are mixed in the same bar. Readers need to repeatedly confirm the position of bar segments to distinguish the **direction of action**, increasing interpretation difficulty;
6. **Unclear expression of specific values**: The stacked chart only allows a rough understanding of the impact of each factor, with no access to precise information;
7. **Poor temporal continuity of factors**: The temporal continuity of each factor is poorly readable, making it difficult to obtain the development trend of individual factors.

3.2 Step 2: Original Chart Reproduction

To ensure the research is based on real data, this project uses Python's **Matplotlib** and **Pandas** tools to highly reproduce the data, color correspondence, and mixed stacking of positive and negative values in the original chart.

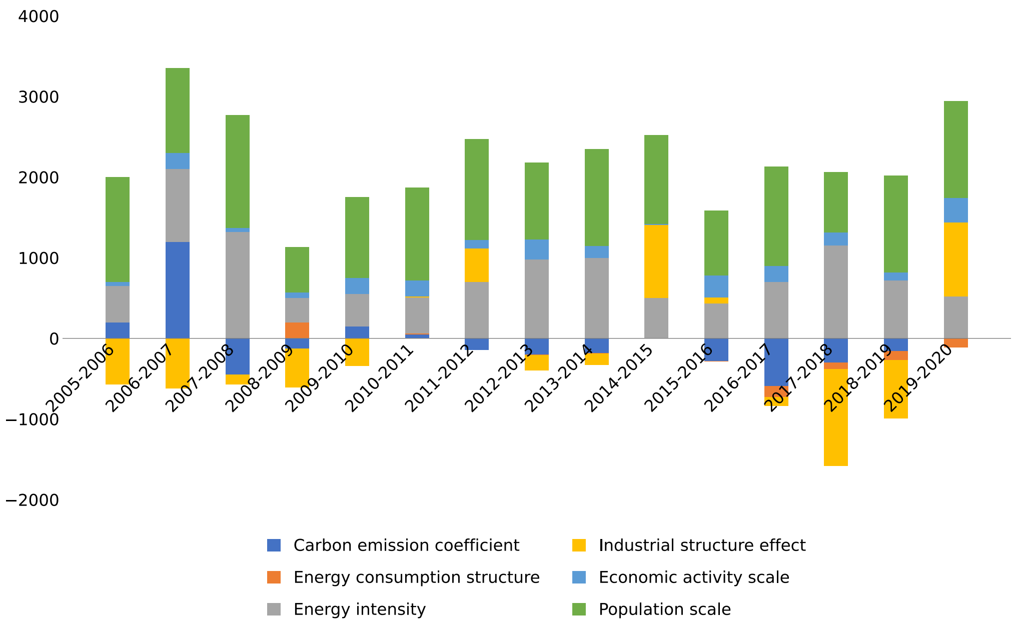

Figure 2: Replicated original chart

3.3 Step 3: Basic Modifications

This stage specifically addresses the core pain points of the original chart—"remote legends, scattered bars, and mixed stacking of positive and negative values"—to establish a clear information transmission framework.

1. **Legend position migration**: Move the bottom legends to the upper right corner to shorten the visual distance from the data, facilitating factor-color comparison, reducing the gaze switching cost between "data and legends", lowering working memory load, and avoiding factor-color confusion;
2. **Adjustment of bar width and spacing**: Widen the bar width to make it larger than the spacing. Visually aggregate the factor modules in the same year to help readers quickly establish an associative understanding of "multiple factors jointly acting on carbon emissions in the same year";
3. **Strengthening the boundary between positive and negative factors**: Add a clear gray reference line at the 0 scale of the vertical axis to separate positive and negative areas. Physically strengthen the

boundary through lines to reduce the difficulty of judging the **direction of factor action** and make identification more intuitive;

4. **Supplementation of title**: Add the main title "Driving Factors of Carbon Emissions in the Yellow River Basin (2005–2020)" to enable readers to quickly obtain the **research object** and **time range**;

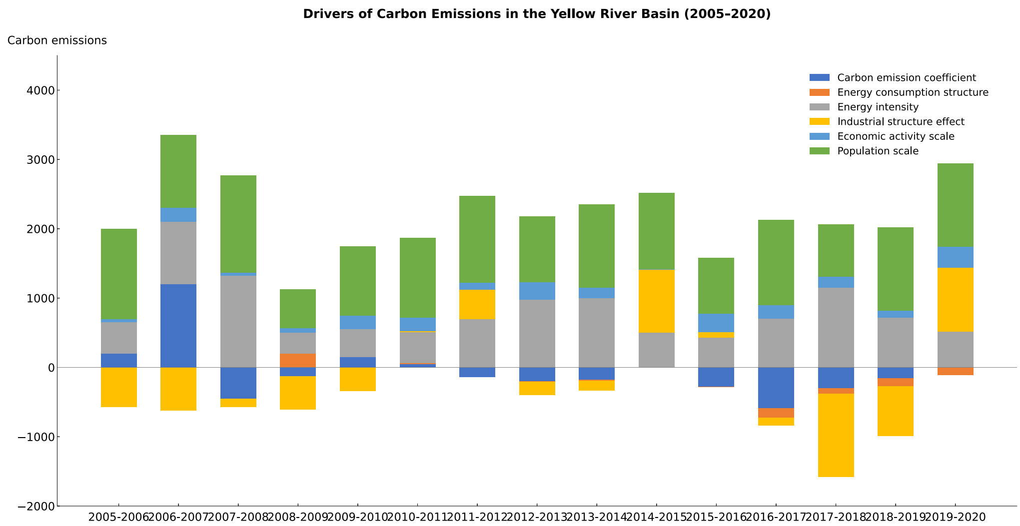

Figure 3: Initial improvement image

3.4 Step 4: Optimization Scheme

Building on the basic modifications, a scatter plot visualization scheme is proposed. Using square-marked scatter plots, it focuses on the "relationship between annual values of individual factors and total annual carbon emissions" to improve **data reading accuracy** and **information dimensions**.

1. **Chart type reconstruction**: Convert the stacked bar chart into a **scatter plot**, using square marks to correspond to "single year-single factor" values. Solve the problem that stacked charts cannot provide **precise readings**, allowing readers to directly obtain values from the coordinate axes without cumulative/split calculations;
2. **Structured design of year intervals**: Assign independent X-axis positions to each year, adjust intervals to avoid mark overlap, fully display all years, and strengthen year boundaries. Ensure that multi-factor marks in a single year are both aggregated and independently readable, balancing overall and local cognition;
3. **Addition of interactive information**: Integrate "year, factor, specific value, and total annual carbon emissions" into hover pop-ups to provide **multi-dimensional data**, reduce visual redundancy, and enrich specific information;
4. **Optimization of visual encoding and reference lines**: Upgrade the y=0 reference line to a thick red line to strengthen the **positive-negative boundary**. Use distinguishable colors to mark different factors and move legends inside the chart to improve **visual comfort** and reduce cognitive fatigue;
5. **Enhancement of trend continuity**: Add connecting lines between data points of each factor to form clear **cross-year trend trajectories**. Strengthen the visual continuity of individual factors over time, facilitating the observation of long-term trends, turning points, and outliers, improving **time series analysis efficiency**, and reducing the cognitive burden of vertical multi-factor comparison.

The research ultimately generates an interactive visualization chart, where readers can obtain multi-dimensional data information through hover interaction:

Click to view the chart of carbon emission drivers

3.5 Step 5: Outcome Publication

To facilitate academic sharing and follow-up research, all complete codes, datasets, and chart outcomes of this project have been uploaded to an open-source **GitHub repository**. The repository includes:

1. **Complete codes**: Provide full-process Python scripts from original chart reproduction and basic modifications to the implementation of optimization schemes;
2. **Original data**: Provide original images and papers on carbon emission driving factors in the Yellow River Basin for reproduction and optimization, ensuring research reproducibility;
3. **Chart outcomes**: Include static images and interactive HTML files of the original chart reproduction, basic modification version, and final optimization scheme;
4. **Documentation**: Provide detailed Markdown documents explaining the optimization logic of each stage to facilitate quick access for other researchers;

GitHub repository address: https://github.com/cccclaip/INFO250_Project2_Group3.git

By publishing as open-source, this study provides reusable technical solutions and theoretical references for practitioners in the field of carbon emission data visualization, helping to improve the social recognition of information visualization technology.

4. Research Findings

The phased optimization successfully addresses the core pain points of the original chart, significantly reducing cognitive load, improving data reading accuracy, and strengthening trend continuity. The optimized chart meets both the needs of precise analysis and rapid information acquisition.

Adhering to data authenticity, the optimized chart achieves a significant improvement in information transmission efficiency and readability. By publishing the research outcomes on GitHub, the core goal of "making high-value data accessible to more people" is realized. The specific values are reflected in the following aspects:

1. **For policy-makers**: Grasp the carbon emission driving logic and transformation trends from multiple dimensions, clarify the core emission reduction status of high-energy-consuming industries such as coal chemical industry and thermal power generation in the secondary industry, increase the promotion of new energy sources such as photovoltaic and wind power, introduce industrial energy efficiency standards, support low-carbon infrastructure construction in regions with rapid population growth, construct an emission reduction system of "industrial regulation + energy substitution + technological upgrading + population adaptation", and promote the transformation of policy-making towards **data-driven**;
2. **For high-energy-consuming enterprises**: Recognize that energy intensity is the main positive driving factor of carbon emissions. Enterprises need to focus on **energy-saving technological upgrading**, increase the proportion of clean energy, layout low-carbon industrial chains, increase investment in low-carbon technology R&D, and respond to policy and market demands;
3. **For the public and student groups**: Understand the carbon emission driving mechanism through visualization, recognize the value of industrial structure optimization and clean energy promotion, understand that economic growth and carbon emissions are not necessarily positively correlated, establish low-carbon concepts, and practice green consumption and low-carbon travel.

5. Division of labor

Name	Tasks
Weixin Kong	improve data visualization and collaboratively revise the report document
Yinuo Sun	reproduce the original graphs and improve data visualization
Shiqi Wang	improve data visualization, organize the code, and write Markdown comments in the code
Junchen Zhou	organize the code, write Markdown comments in the code, and compile the report document

6. Conclusion and Future Direction

This study follows a 'Replicate - Diagnose - Optimize' technical approach to enhance data readability. In the era of **data explosion**, visualization is not a simple listing and presentation of data, but a key link connecting data and users. Its core lies in, on the premise of adhering to data authenticity, deeply aligning with the **cognitive laws** of audiences with different knowledge backgrounds, and realizing the goals of reducing users' cognitive load and improving information transmission efficiency through design means such as optimizing visual layout and strengthening information differentiation.

Producing high-quality visualization charts requires accurately positioning the core value and application scenarios of data, matching the actual needs and cognitive levels of target users, and ultimately completing the efficient transmission of information in a presentation form that conforms to cognitive laws.

Future research can focus on the deepening and expansion of visualization in professional fields, and develop more targeted chart types and interactive modes for the characteristics of **multi-source heterogeneous data** in subdivided scenarios such as industrial structure inclination.