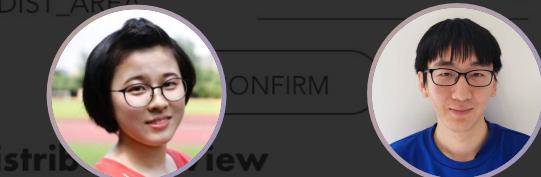


# Interactive Visual Exploration of Longitudinal Historical Career Mobility Data



Yifang Wang<sup>1,2,3</sup> Hongye Liang<sup>2,3</sup>



Zikun Deng<sup>2,3</sup> Cameron Campbell<sup>1</sup>



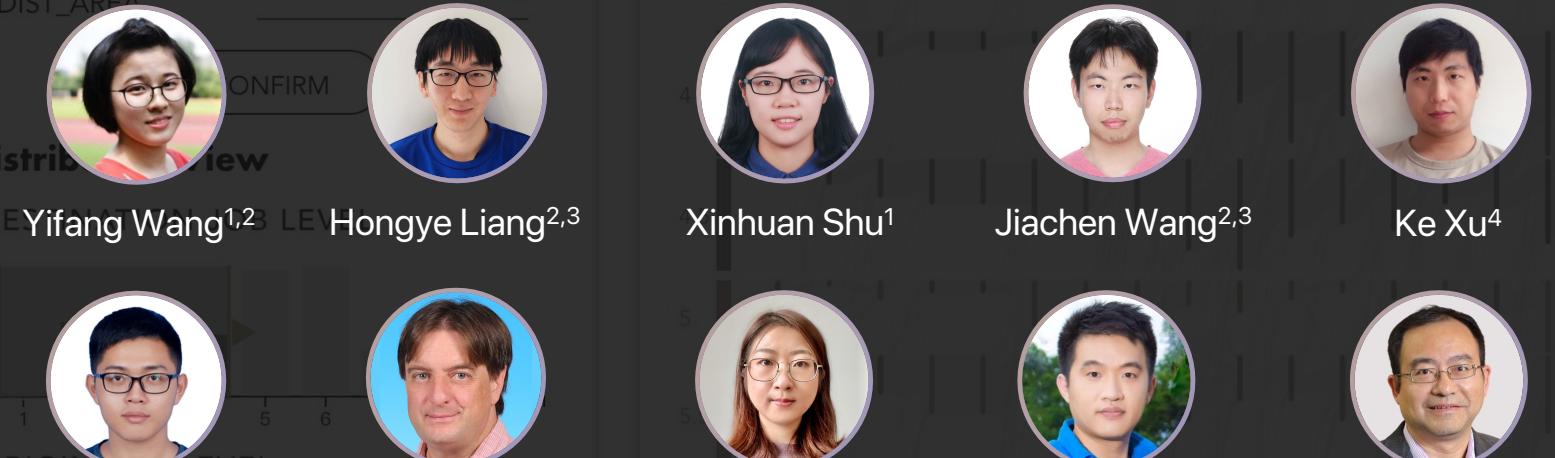
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY



NYU RENMIN UNIVERSITY OF CHINA



VIS 2021



Xinhuan Shu<sup>1</sup>

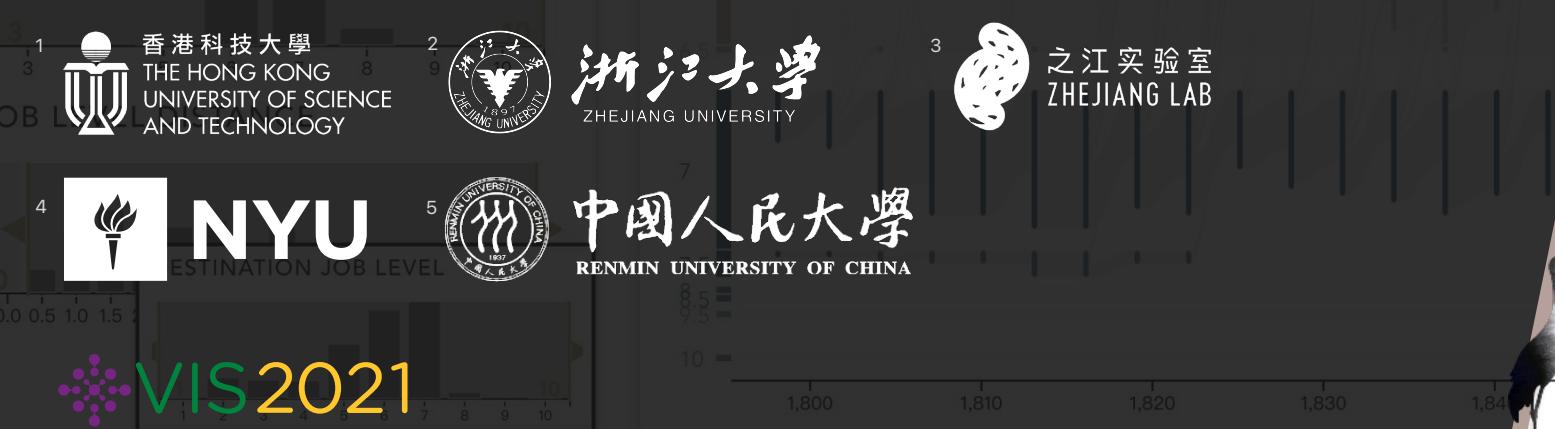
Jiachen Wang<sup>2,3</sup>

Ke Xu<sup>4</sup>

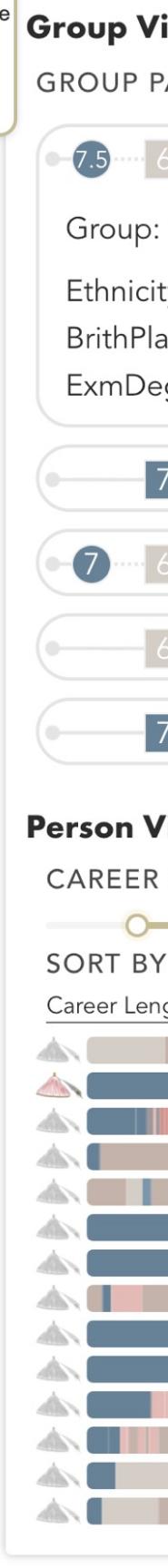
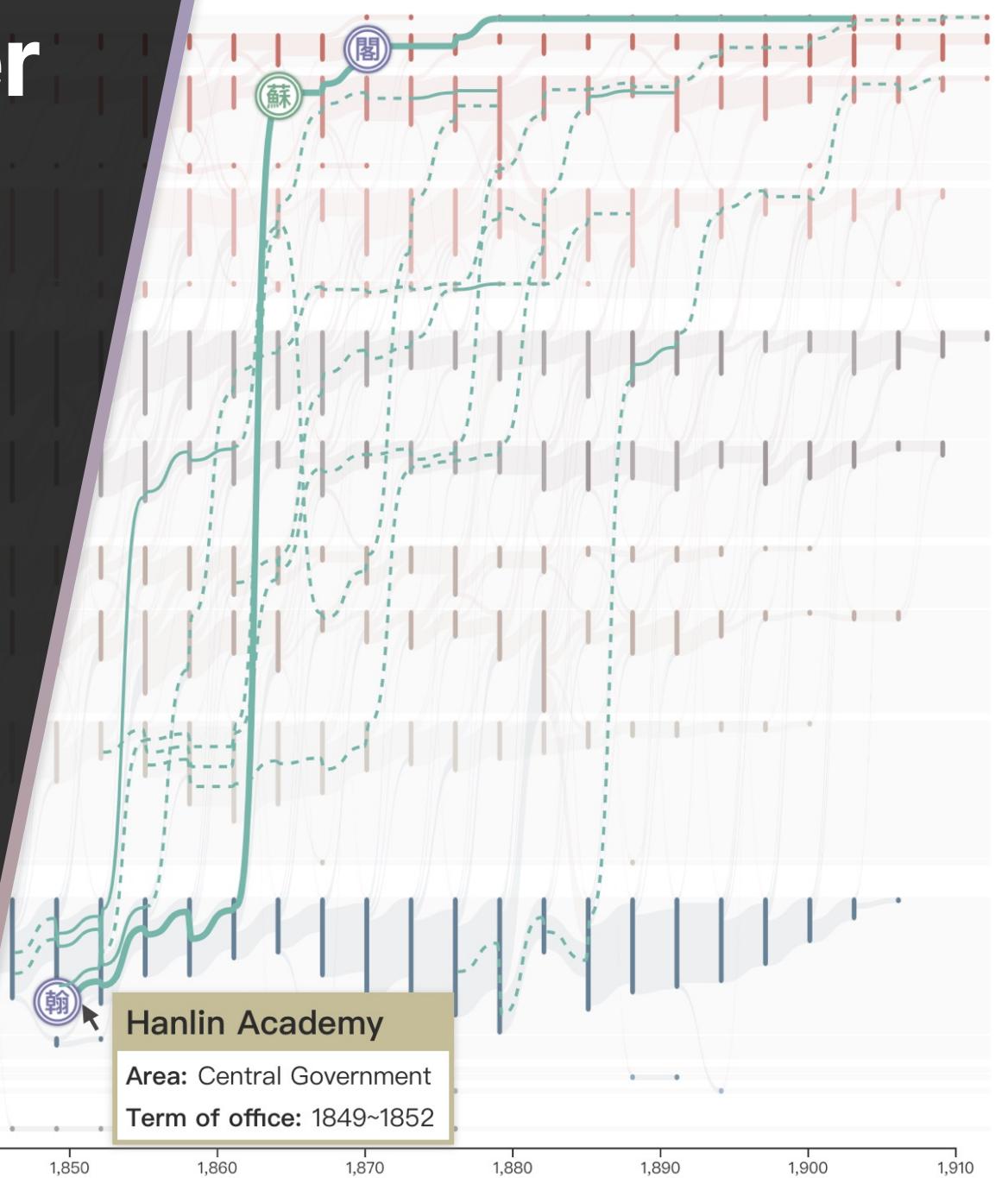
Bijia Chen<sup>5</sup>

Yingcai Wu<sup>2,3</sup>

Huamin Qu<sup>1</sup>



ZHEJIANG UNIVERSITY  
ZHEJIANG LAB



# Project Background



Quantitative Historical Datasets  
CGED-Q  
Career Mobility

- Digital Humanities
- History
- Sociology
- Demography
- ...
- Career
- Health
- Family
- Migration
- ...



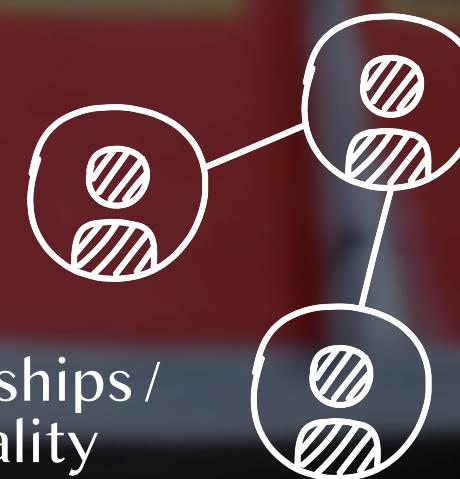
# Historical Quantitative Datasets

# China Government Employee Database-Qing (CGED-Q)



Social Relationships /  
Social Inequality

- One-year project
- CGED-Q records the career trajectories of over 340,000 government officials in the bureaucracy of Qing China from 1760 to 1912
- Career Mobility: the study of career trajectories and the factors influencing them.



# Career Mobility Analysis

Hout, 1983

Father's Occupation	Son's Occupation						Total
	Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm	Total	
<b>Inflow Percentages</b>							
Upper Nonmanual	34.5	17.6	11.5	8.1	1.8	14.7	
Lower Nonmanual	17.7	17.7	9.7	8.8	2.1	11.3	
Upper Manual	19.5	21.9	32.6	21.1	4.8	20.5	
Lower Manual	18.4	30.8	29.4	41.8	10.5	30.1	
Farm	10.0	12.0	16.8	20.2	80.9	23.4	
Total	100.0	100.0	100.0	100.0	100.0	100.0	

Jarvis and Song, 2017

Model	df	$L^2$	p-value	BIC	$\Delta$
<i>Men (N = 56,200)</i>					
<i>Baseline Models</i>					
0a: Origin + Destination + Period	27,972	244,675	<.01	-61,245	69.0
1a: 0a + Origin $\times$ Period + Destination $\times$ Period	27,380	239,997	<.01	-59,449	68.5
2a: 1a + Parameter-ABCD	27,292	27,431	.28	-271,052	18.2
<i>Period-Varying Topological Mobility Models</i>					
3a: 2a + Period $\times$ Parameter-A	26,992	24,874	>.99	-270,328	14.2
4a: 2a + Period $\times$ Parameter-AB	26,956	24,723	>.99	-270,086	14.1
5a: 2a + Period $\times$ Parameter-ABC	26,944	24,661	>.99	-270,017	14.1
6a: 2a + Period $\times$ Parameter-ABCD	26,940	24,648	>.99	-269,986	14.1
7a: 2a + Period $\times$ Parameter-A'B'C'D'	27,276	25,659	>.99	-272,650	16.0
<i>Women (N = 47,180)</i>					
<i>Baseline Models</i>					
0b: Origin + Destination + Period	27,972	174,593	<.01	-126,434	66.6
1b: 0b + Origin $\times$ Period + Destination $\times$ Period	27,380	166,191	<.01	-128,465	65.7
2b: 1b + Parameter-ABCD	27,292	19,635	>.99	-274,074	15.2
<i>Period-Varying Topological Mobility Models</i>					
3b: 2b + Period $\times$ Parameter-A	26,992	18,287	>.99	-272,193	12.3
4b: 2b + Period $\times$ Parameter-AB	26,956	18,154	>.99	-271,939	12.2
5b: 2b + Period $\times$ Parameter-ABC	26,944	18,102	>.99	-271,862	12.1
6b: 2b + Period $\times$ Parameter-ABCD	26,940	18,077	>.99	-271,844	12.1

- Social scientists
- Group-level analysis

# Career Mobility Analysis

- Hypothesis-driven and ill-suited for exploration

- Social scientists  
Group-level analysis

# Career Mobility Analysis

Hout, 1983

Father's Occupation	Son's Occupation						Total
	Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm	Total	
<b>Inflow Percentages</b>							
Upper Nonmanual	34.5	17.6	11.5	8.1	1.8	14.7	
Lower Nonmanual	17.7	17.7	9.7	8.8	2.1	11.3	
Upper Manual	19.5	21.9	32.6	21.1	4.8	20.5	
Lower Manual	18.4	30.8	29.4	41.8	10.5	30.1	
Farm	19.0	12.0	16.8	20.2	80.9	23.4	
Total	100.0	100.0	100.0	100.0	100.0	100.0	

- Hypothesis-driven and ill-suited for exploration

See Origin > Destination > Period	St. 002	St. 400	St. 800	St. 1200	St. 2000
See See > OriginPeriod > DestinationPeriod	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period > ABCD	St. 000	St. 400	St. 800	St. 1200	St. 2000
Period Mobility Theoretical Mobility Models	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period A	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period B	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period C	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period D	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period ABCD	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Periods Period ABCD	St. 000	St. 400	St. 800	St. 1200	St. 2000
Women (N = 47,180)	St. 000	St. 400	St. 800	St. 1200	St. 2000
Random Models	St. 000	St. 400	St. 800	St. 1200	St. 2000
See Origin > Destination > Period	St. 002	St. 400	St. 800	St. 1200	St. 2000
See See > OriginPeriod > DestinationPeriod	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period > ABCD	St. 000	St. 400	St. 800	St. 1200	St. 2000
Period Mobility Theoretical Mobility Models	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period A	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period B	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period C	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period D	St. 000	St. 400	St. 800	St. 1200	St. 2000
See See > Period ABCD	St. 000	St. 400	St. 800	St. 1200	St. 2000

Qu et al., 2016

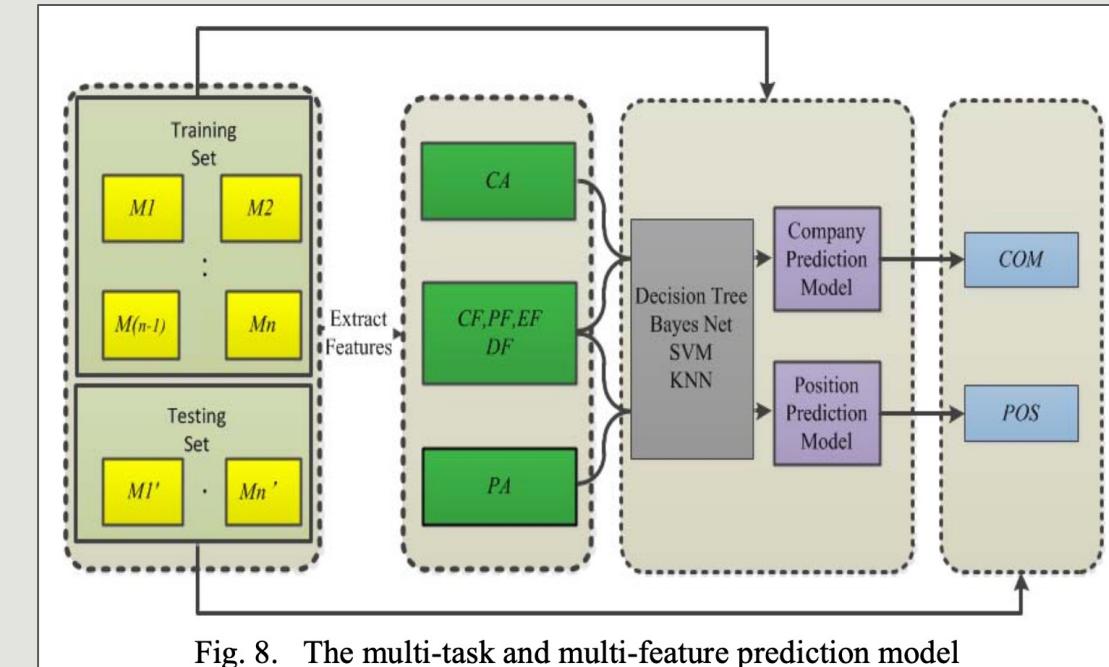
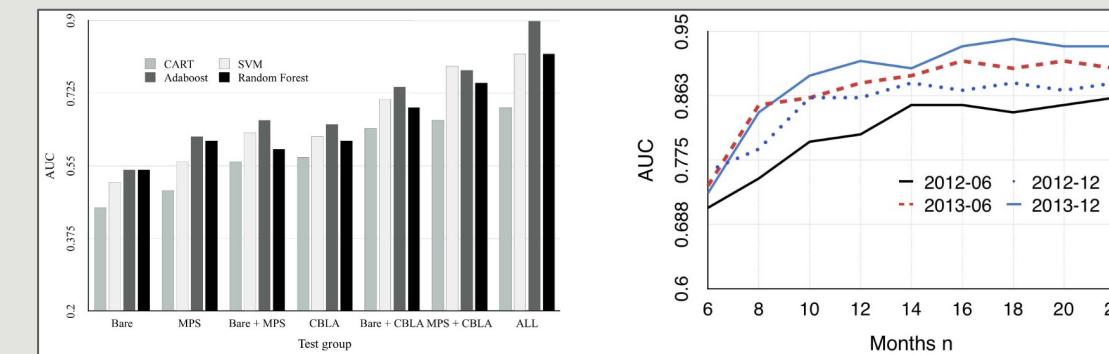


Fig. 8. The multi-task and multi-feature prediction model

Xu et al., 2015



- Social scientists  
Group-level analysis

- Data scientists  
Individual-level analysis

# Career Mobility Analysis

Hout, 1983

Father's Occupation	Son's Occupation						Total
	Upper Nonmanual	Lower Nonmanual	Upper Manual	Lower Manual	Farm	Total	
<b>Inflow Percentages</b>							
Upper Nonmanual	34.3	17.6	11.5	8.1	1.8	14.7	
Lower Nonmanual	17.7	17.7	9.7	8.8	2.1	11.3	
Upper Manual	19.5	21.9	32.6	21.1	4.8	20.5	
Lower Manual	18.4	30.8	29.4	41.8	10.5	30.1	
Farm	19.0	12.0	16.8	20.2	80.9	23.4	
Total	100.0	100.0	100.0	100.0	100.0	100.0	

- Hypothesis-driven and ill-suited for exploration

	1962	1964	1966	1968	1970	1972	1974	1976	1978	1980	1982	1984	1986	1988	1990	1992	1994	1996	1998	2000	2002	2004	2006	2008	2010	2012	2014	2016	2018	2020						
ln(Odds = Distribution + Period)	27.972	294.407	-0.001	-461.246	00.0																															
ln(Odds = Distribution*Period)	27.999	299.397	-0.001	-466.449	00.0																															
ln(Odds = Persistence+ABC)	27.999	27.401	0.001	-471.002	00.2																															
Period-Varying Theoretical Mobility Models																																				
ln(Odds = Persistence-A)	26.992	26.384	-0.001	-476.008	04.2																															
ln(Odds = Persistence-B)	26.999	26.782	-0.001	-476.008	04.2																															
ln(Odds = Persistence-ABC)	26.994	26.983	-0.001	-476.017	04.2																															
ln(Odds = Persistence-ABC2)	26.999	26.988	-0.001	-486.008	04.2																															
ln(Odds = Persistence-Persistence ABC/2)	27.279	25.808	-0.001	-473.000	04.2																															
Women (N = 47,180)																																				
Baseline Models																																				
ln(Odds = Distribution + Period)	27.972	294.407	-0.001	-461.246	00.0																															
ln(Odds = Distribution*Period)	27.999	299.397	-0.001	-466.449	00.0																															
ln(Odds = Persistence+ABC)	27.999	27.401	0.001	-471.002	00.2																															
Period-Varying Theoretical Mobility Models																																				
ln(Odds = Persistence-A)	26.992	26.384	-0.001	-476.008	04.2																															
ln(Odds = Persistence-B)	26.999	26.782	-0.001	-476.008	04.2																															
ln(Odds = Persistence-ABC)	26.994	26.983	-0.001	-476.017	04.2																															
ln(Odds = Persistence-ABC2)	26.999	26.988	-0.001	-486.008	04.2																															

- Social scientists  
Group-level analysis

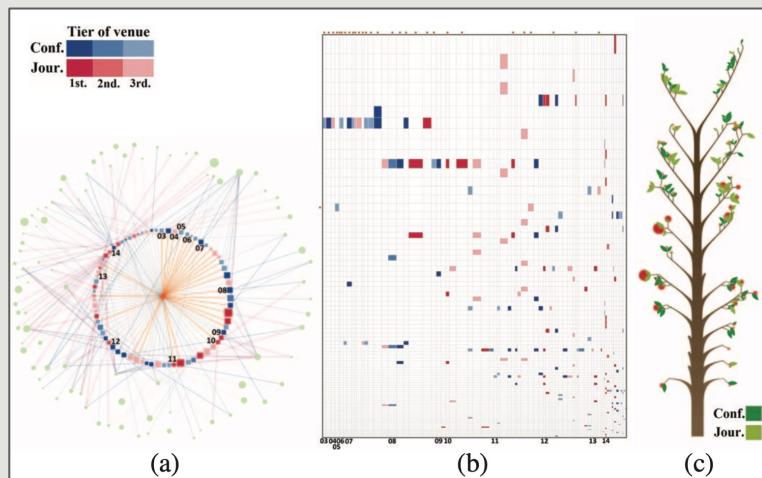
Qu et al., 2016



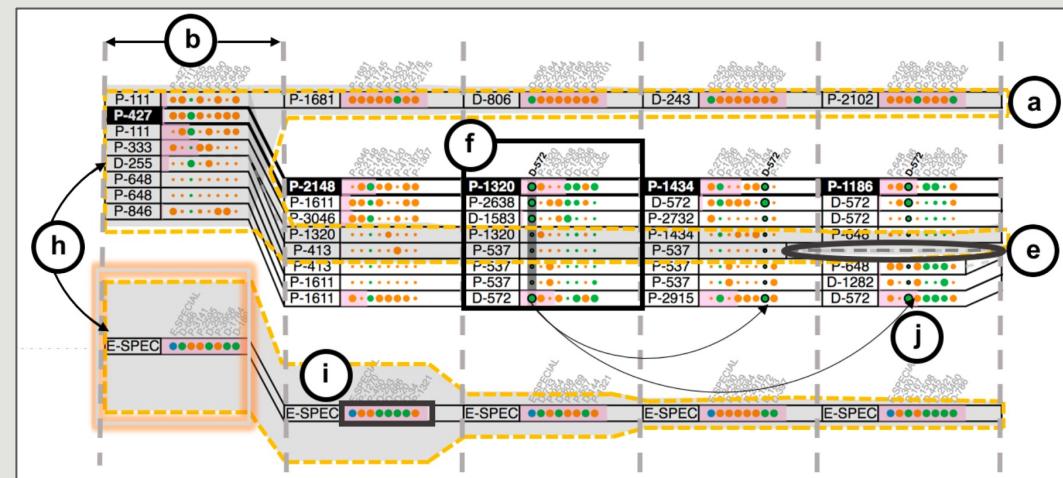
- Data scientists  
Individual-level analysis

# Career Data Visualization (1/2)

Fung et al., 2016



Guo et al., 2018



Du et al., 2016



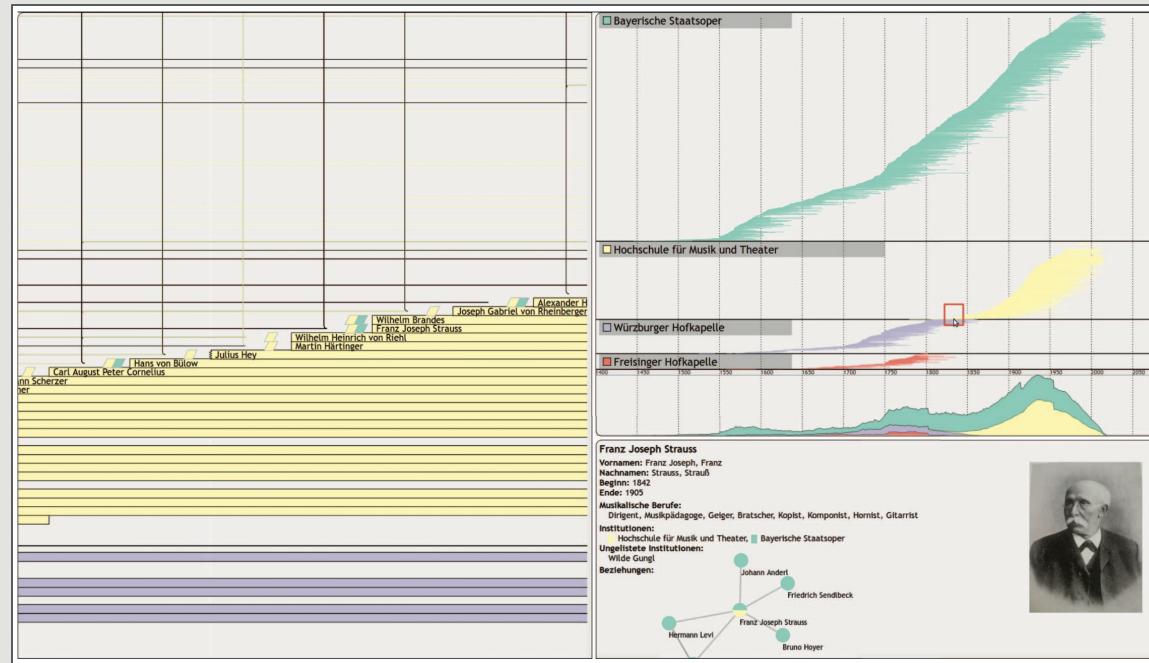
- Network summarization

- Sequence summarization

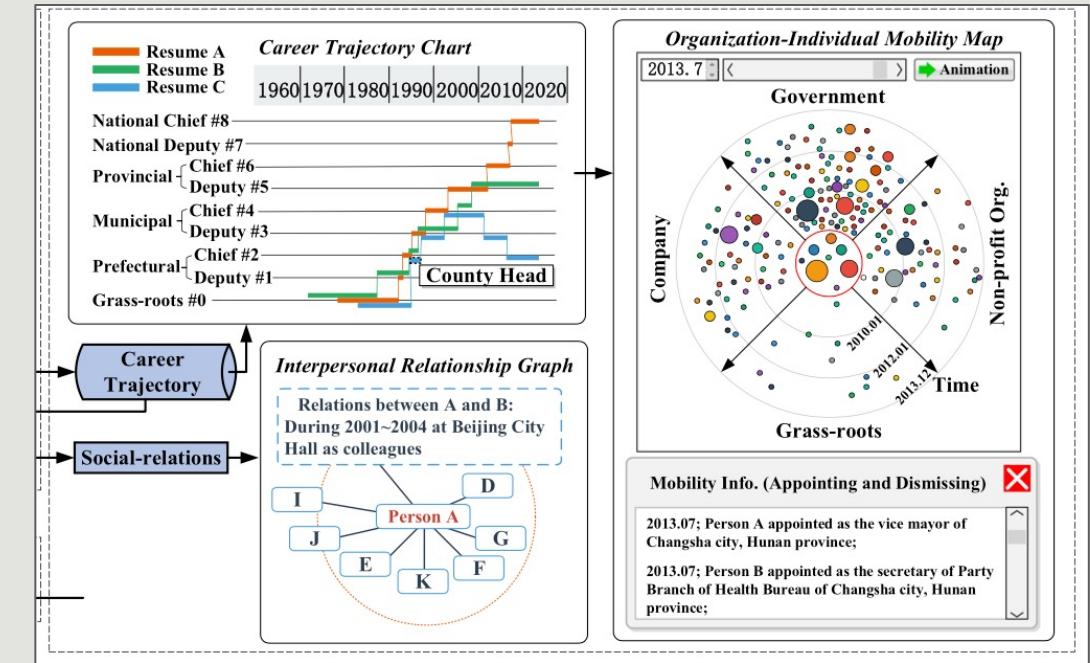
- Similarity comparison

# Career Data Visualization (2/2)

Khulusi et al., 2019



Zhang and Wang, 2019



- Multi-task analysis

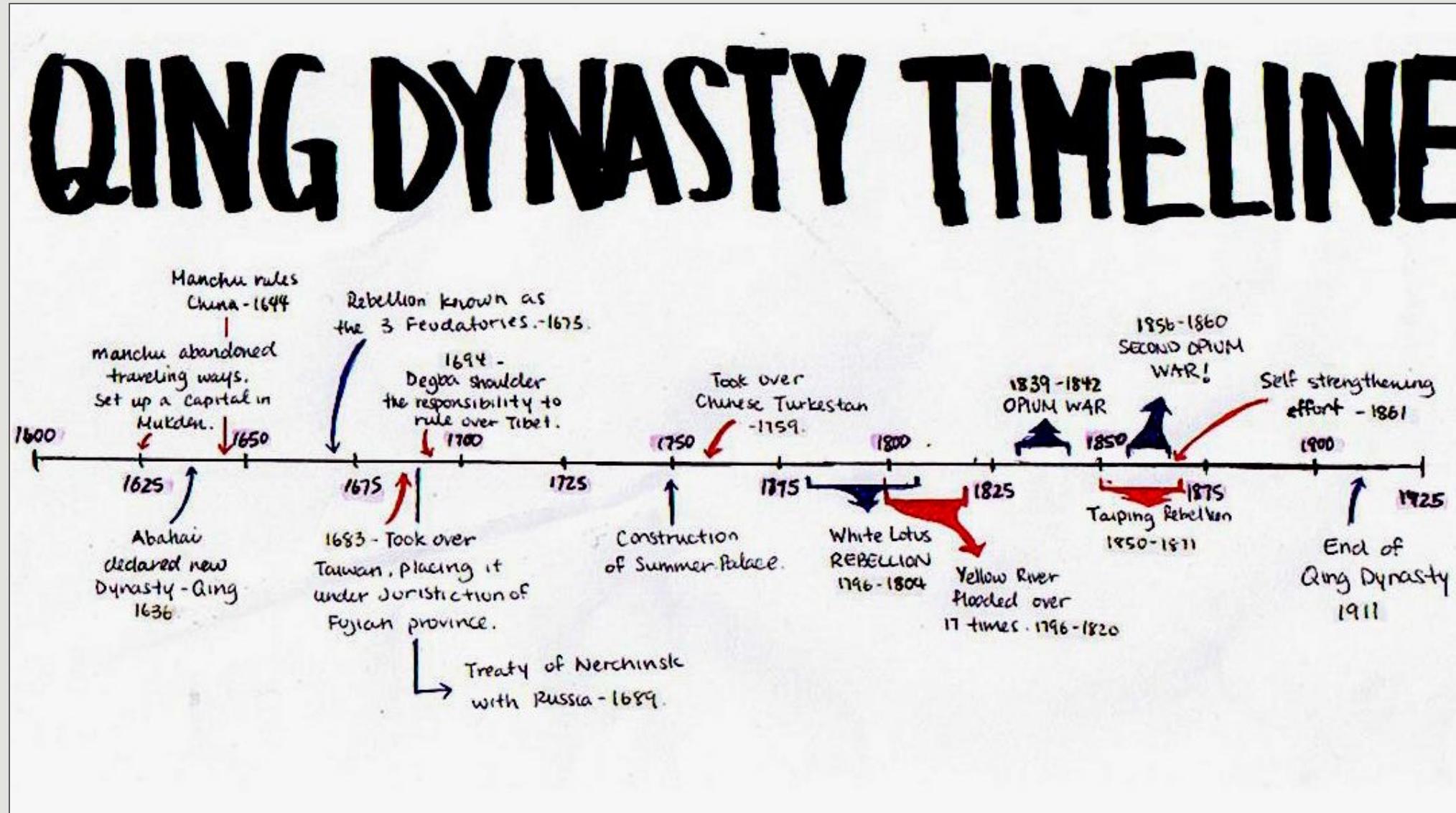
# Career Data Visualization (2/2)

- Support a limited number of career comparisons and lose the overall context
  - Are limited by the short time range



- Multi-task analysis

# Data Description



# Data Description

- **Timestamp:** The year and season covered by the record
- **Name:** The official's real name in the Qing dynasty
- **Unique ID:** A 12-character unique identifier of each official generated by experts

Personal Info

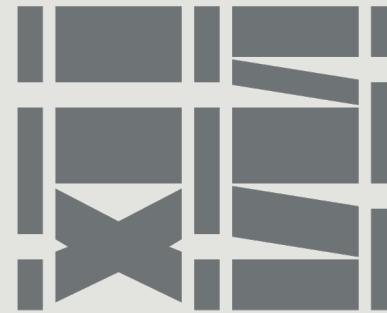
Job Info

# Data Description

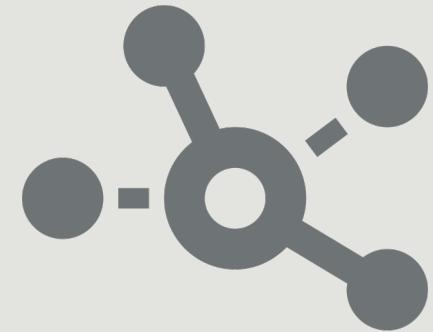
- **Timestamp:** The year and season covered by the record
- **Name:** The official's real name in the Qing dynasty
- **Unique ID:** A 12-character unique identifier of each official generated by experts
- **Birthplace:** The geographic origin of the official
- **Family Background:** A identity indicating whether the official was associated with the imperial lineage
- **Ethnicity:** Three types of officials are identified based on ethnicities: Manchu, Mongol, or Han
- **Exam Degree (科举结果):** The examination or purchased degree held by the official. Those with high examination degrees were political elites
- **Job Location:** The geographical location of the official's current job
- **Job Department:** The department in the bureaucracy where the official works. We classified them into fifteen categories according to experts' suggestions
- **Job Level (品级):** The administrative rank of the job in the bureaucratic hierarchy, represented by a number (ranging from 10 to 1 with 0.5 as a step).

# Challenges (1/3)

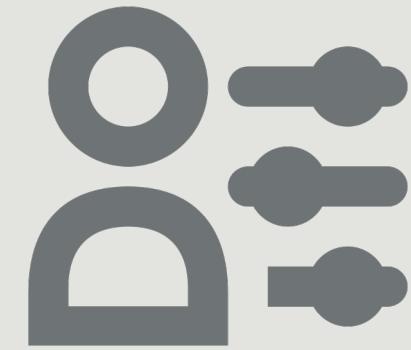
- How to visualize a large volume of longitudinal career data with a complex data structure?



- Temporal
- 1760 ~ 1912



- Network
  - Colleagues
  - Townsmen
  - Classmates



- Multi-attributes
  - Personal Info
  - Job Info

# Challenges (2/3)

- How to visualize a large volume of longitudinal career data with a complex data structure?
- How to extract and highlight social groups and social relationships from this large dataset?



- Predefined Group vs. Latent Group



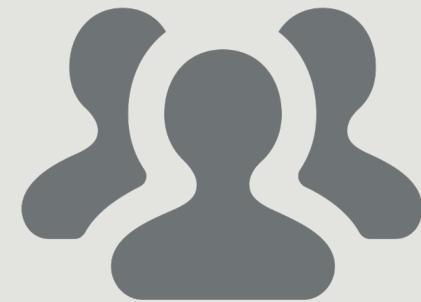
- Semi-automatic

# Challenges (3/3)

- How to visualize a large volume of longitudinal career data with a complex data structure?
- How to extract and highlight social groups and social relationships from this large dataset?
- How to support multi-level mobility analysis and reasoning?



- Overall Level



- Group Level



- Individual Level

# Tasks and System Design



Tasks  
Glyph  
Population Flow

# Experts and Tasks

- Experts



中國人民大學清史研究所  
THE INSTITUTE OF QING HISTORY

Overall



- What are the general characteristics of career mobility?
- What special features do the groups with vertical movements have at different time periods?

Group



- What are the characteristics of different social groups?
- What is the mobility pattern for each group?

Individual

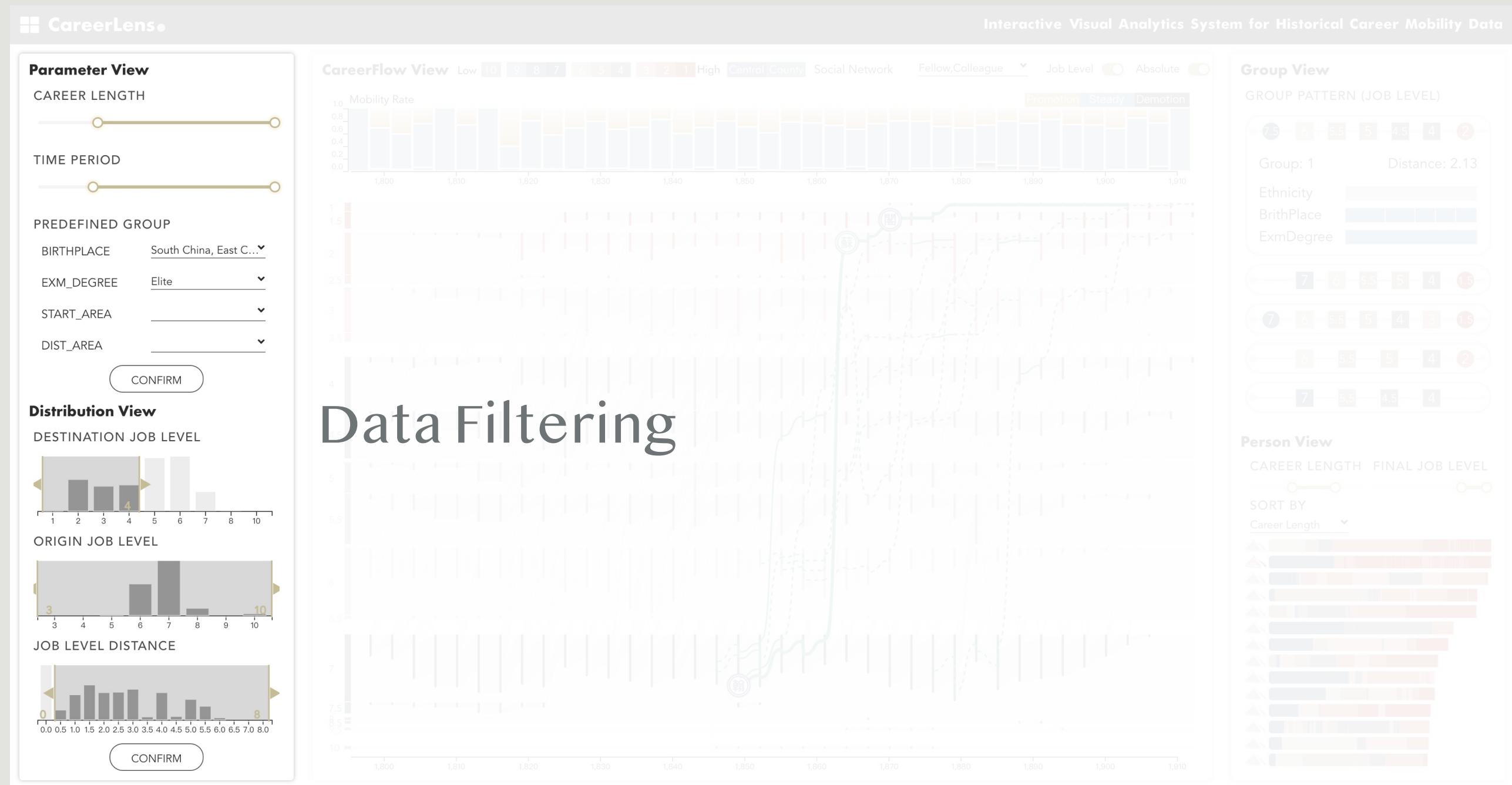


- What are the mobility characteristics for different individuals?
- How do the mobility patterns of each individual and his social relationships change over time?

# CareerLens



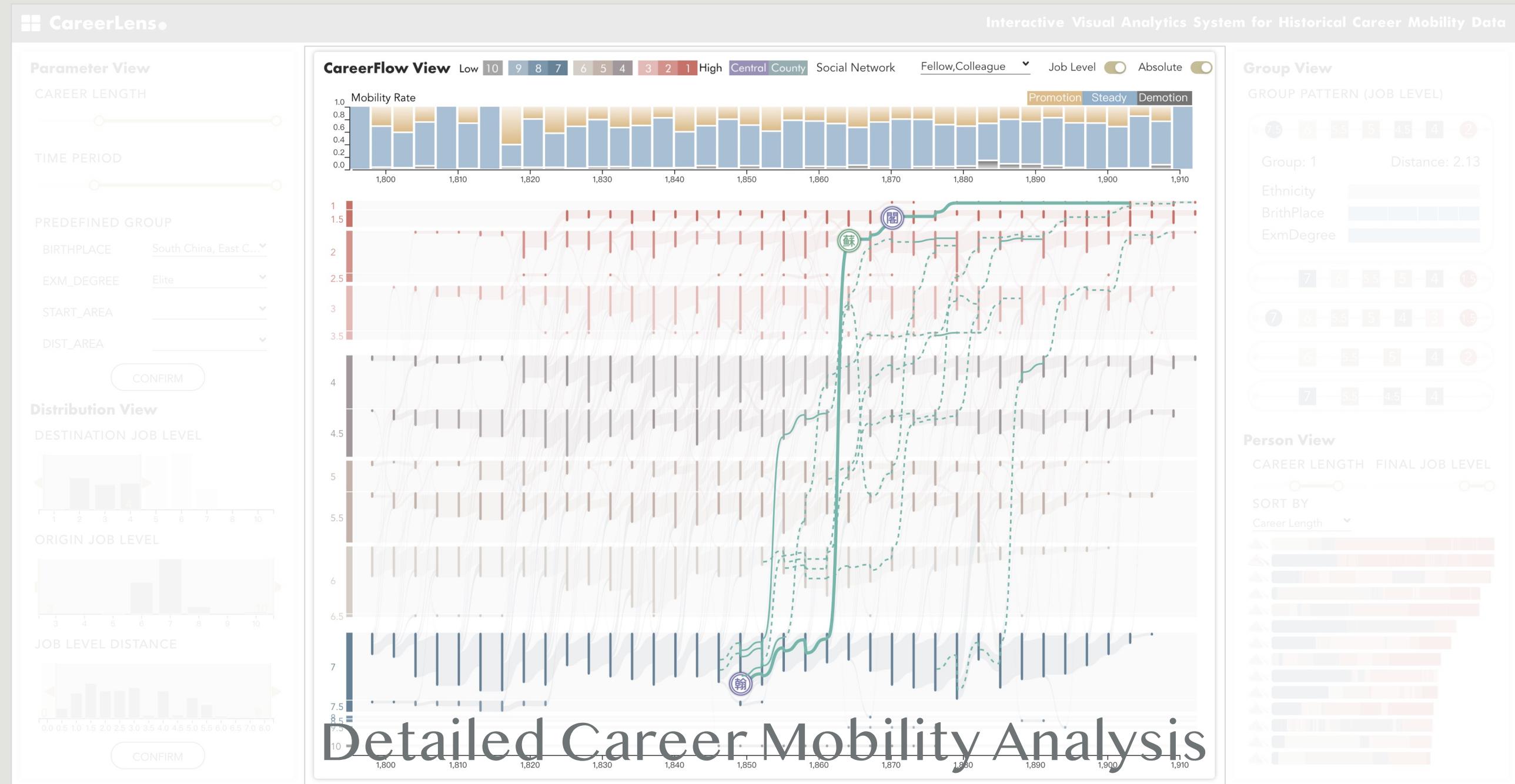
# Visual Design



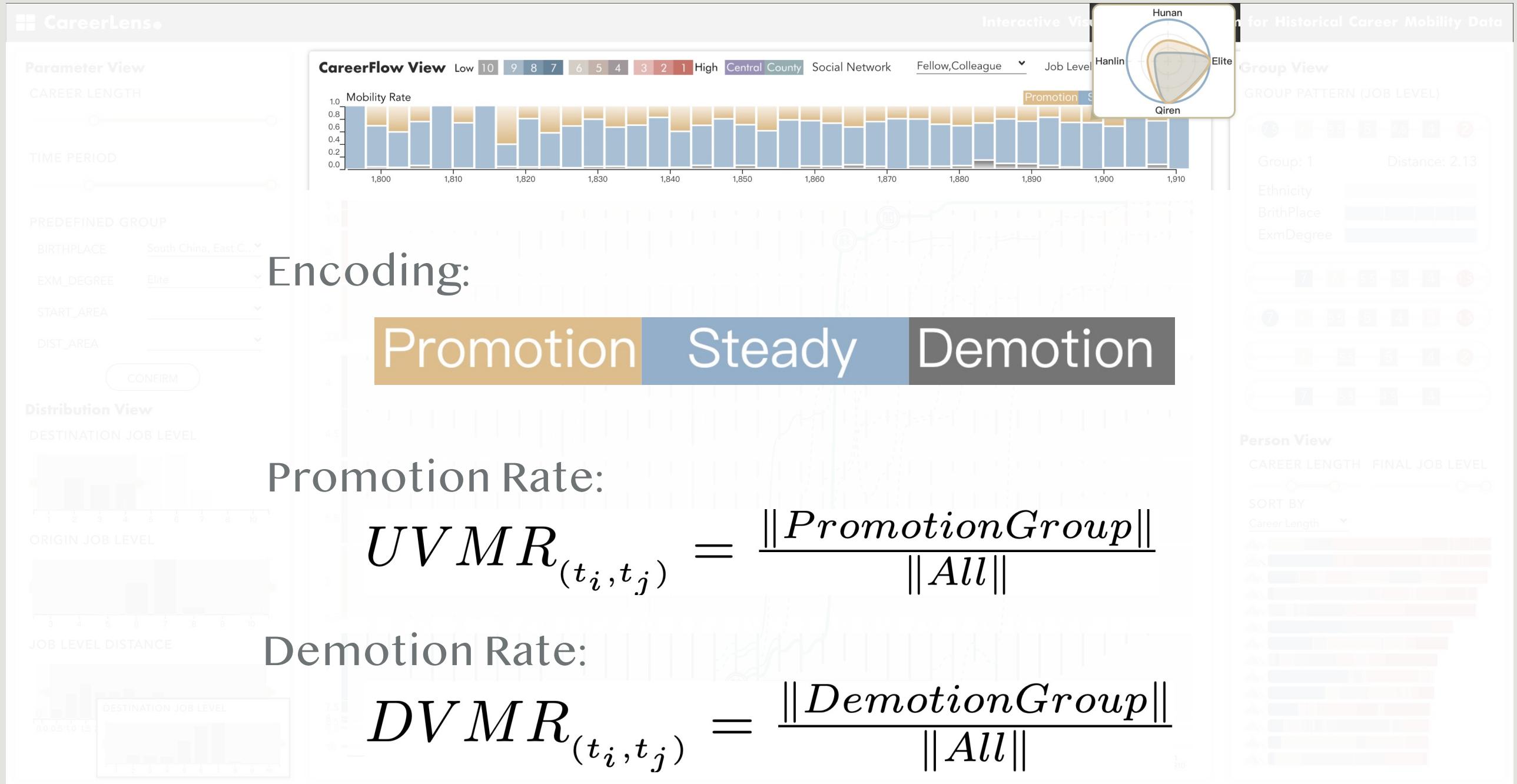
# Visual Design



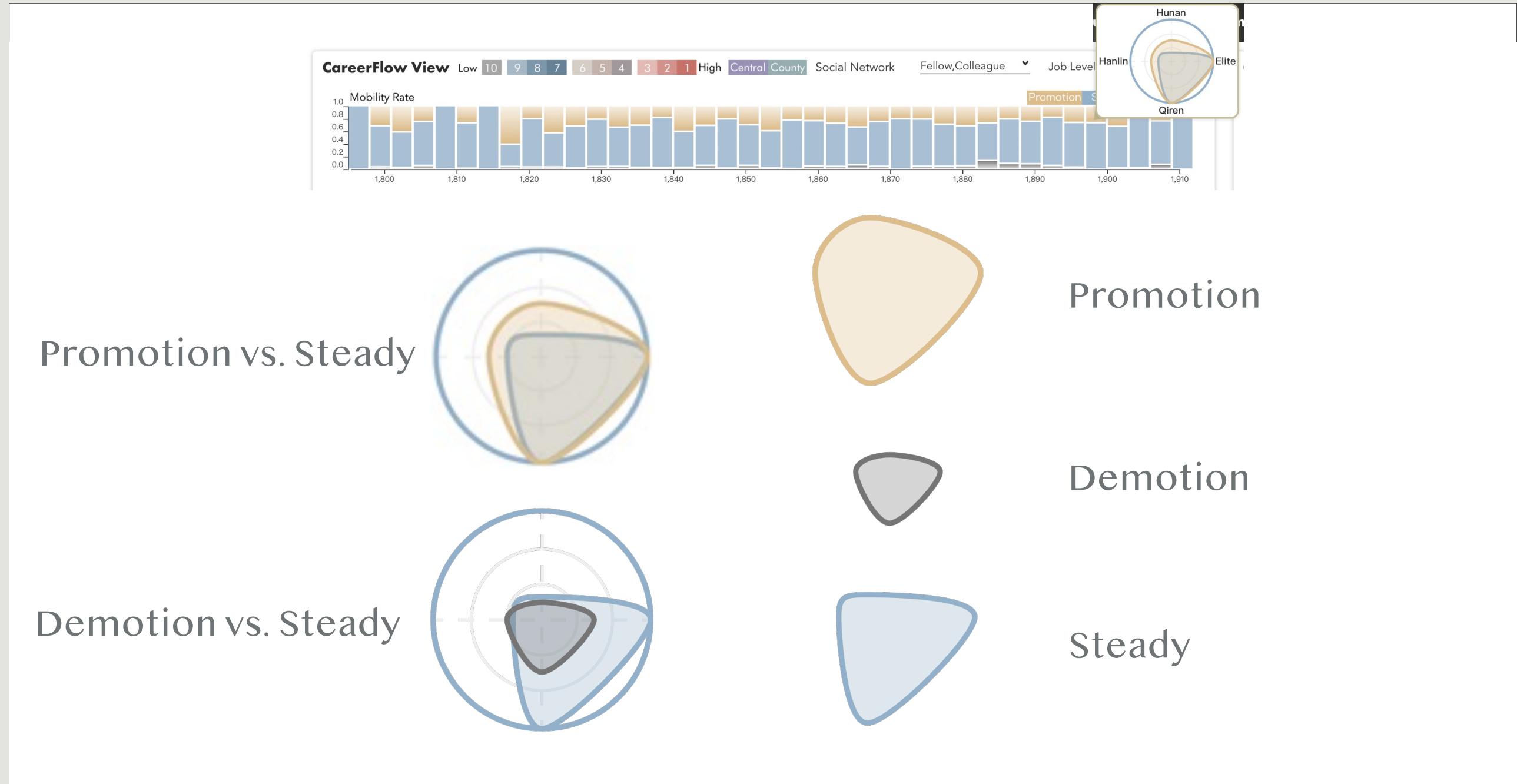
# Visual Design



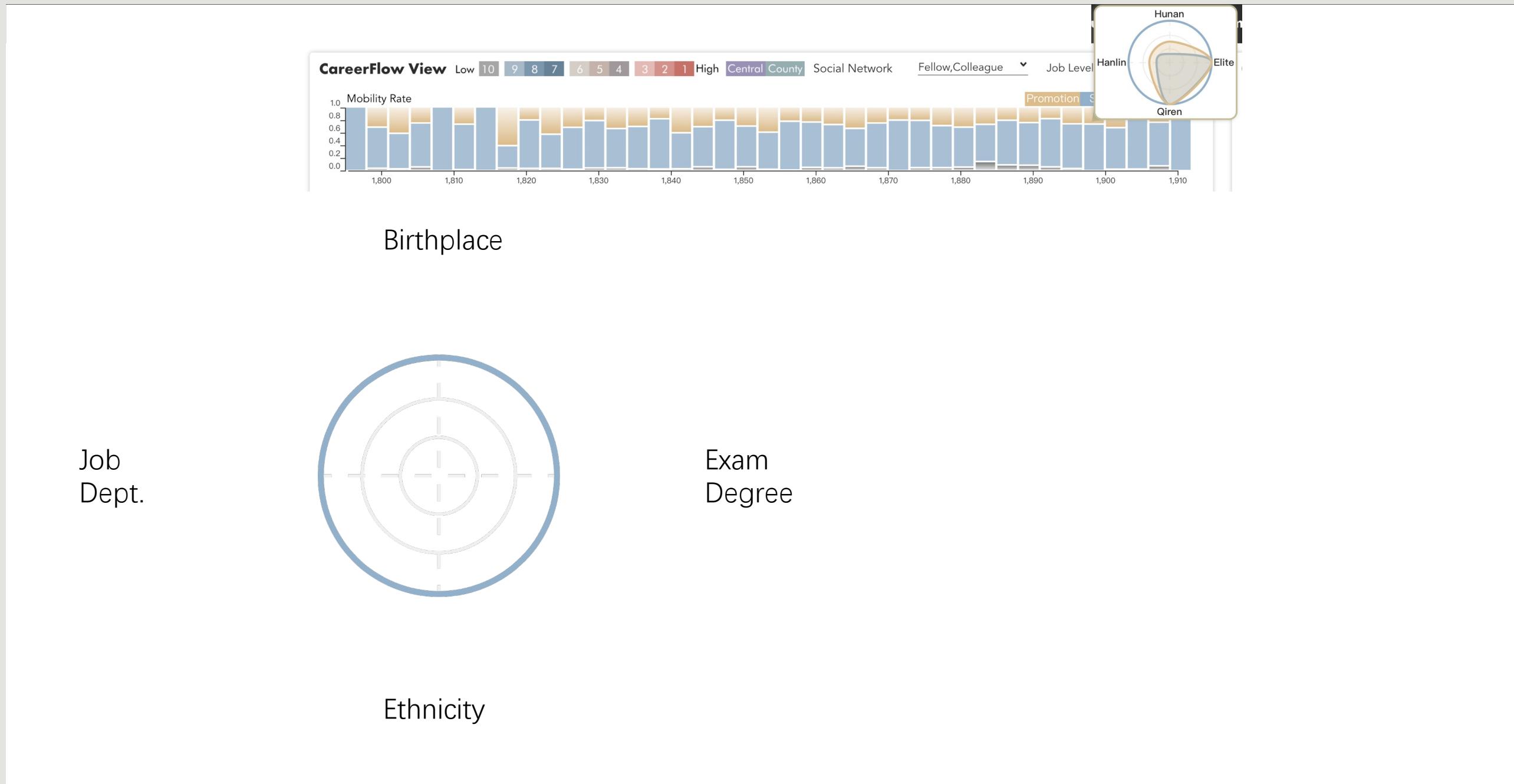
# Visual Design



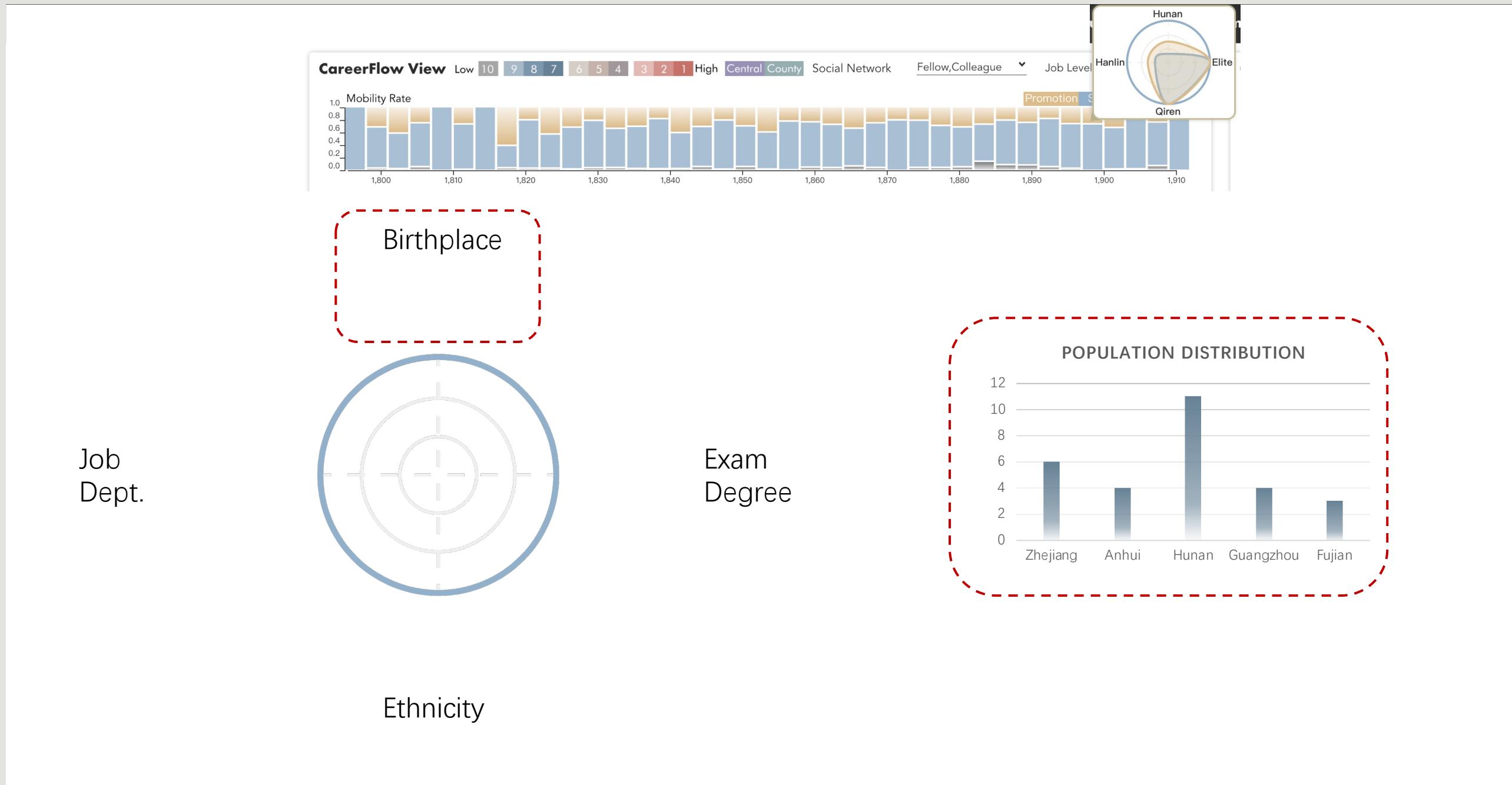
# Visual Design



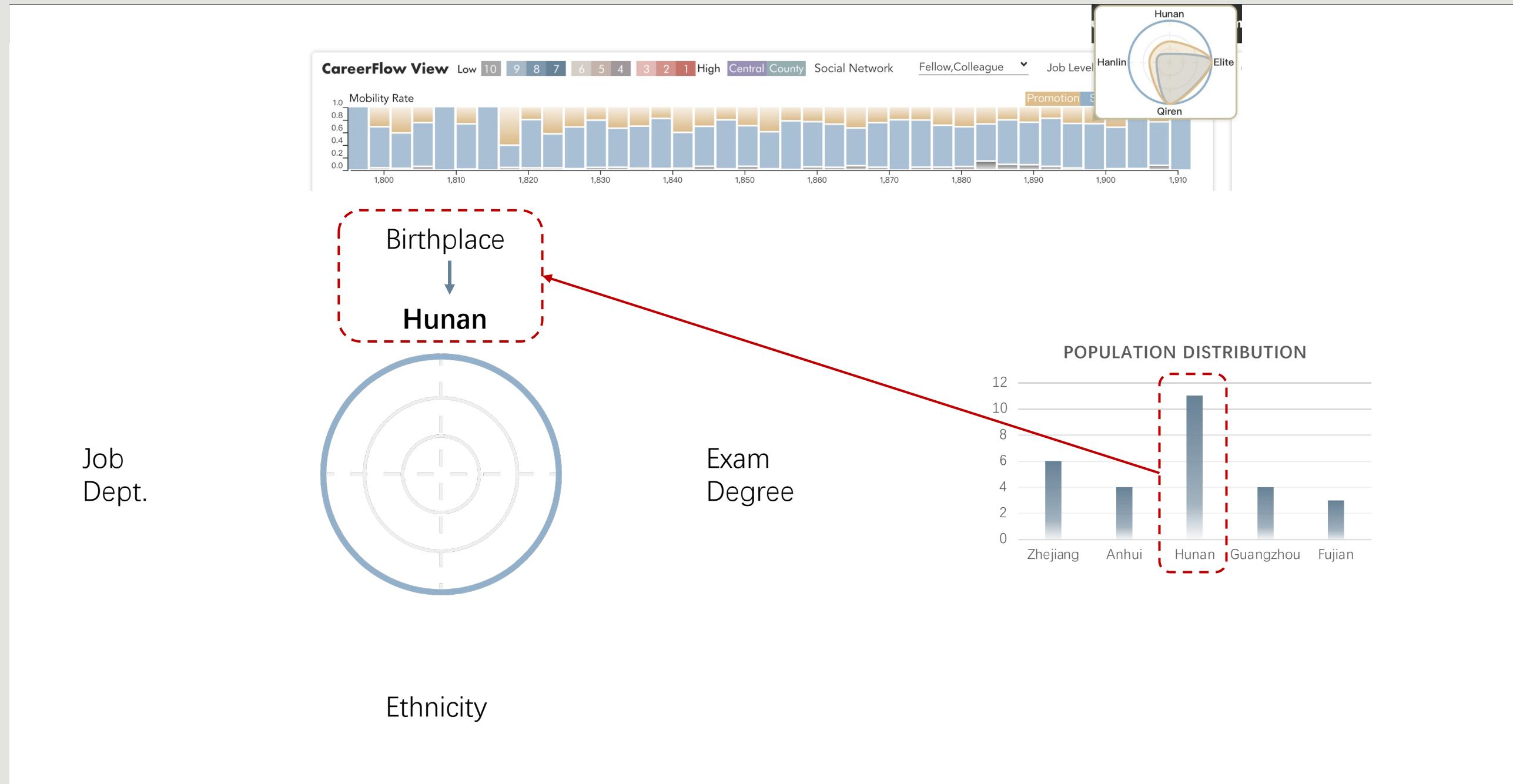
# Visual Design



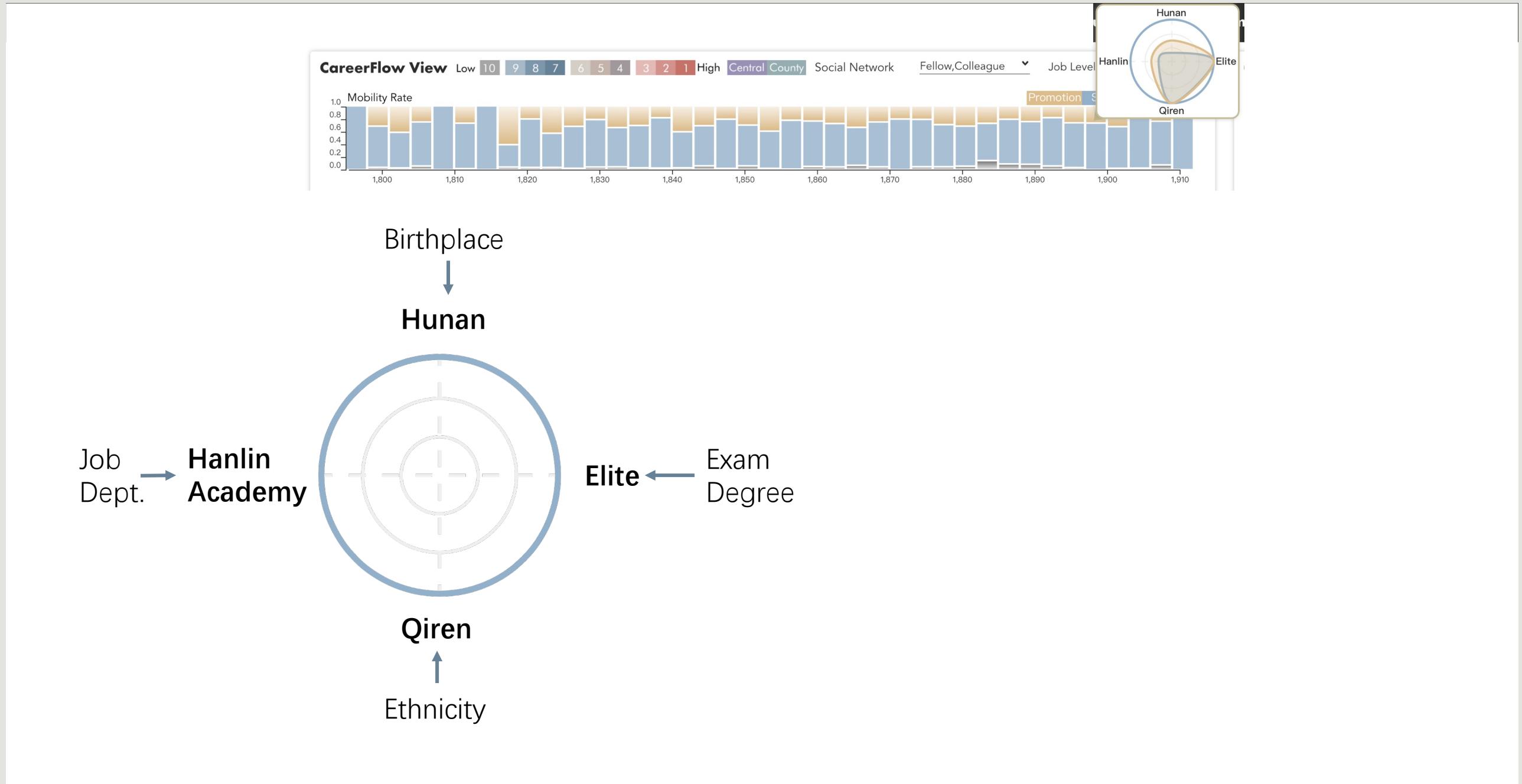
# Visual Design



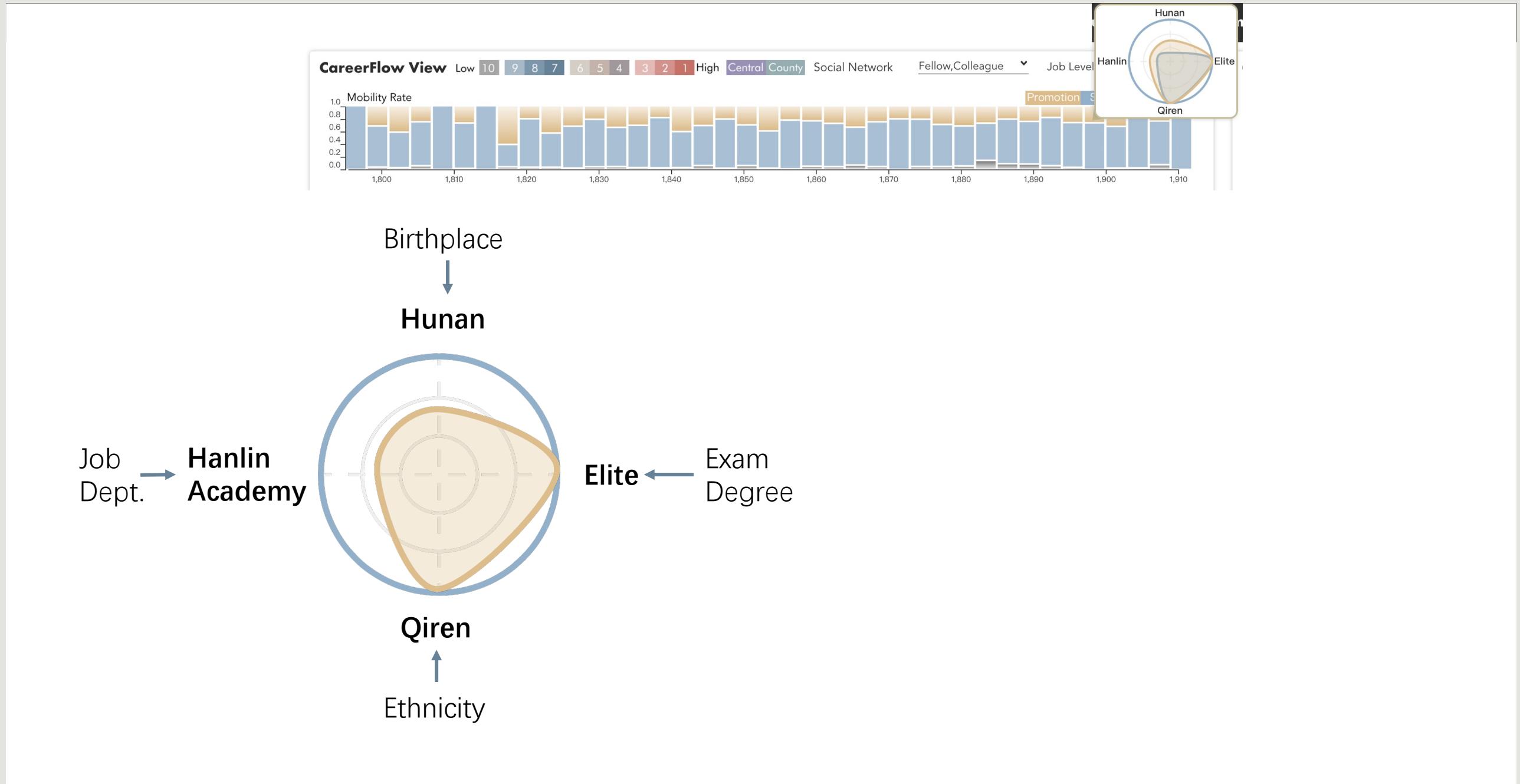
# Visual Design



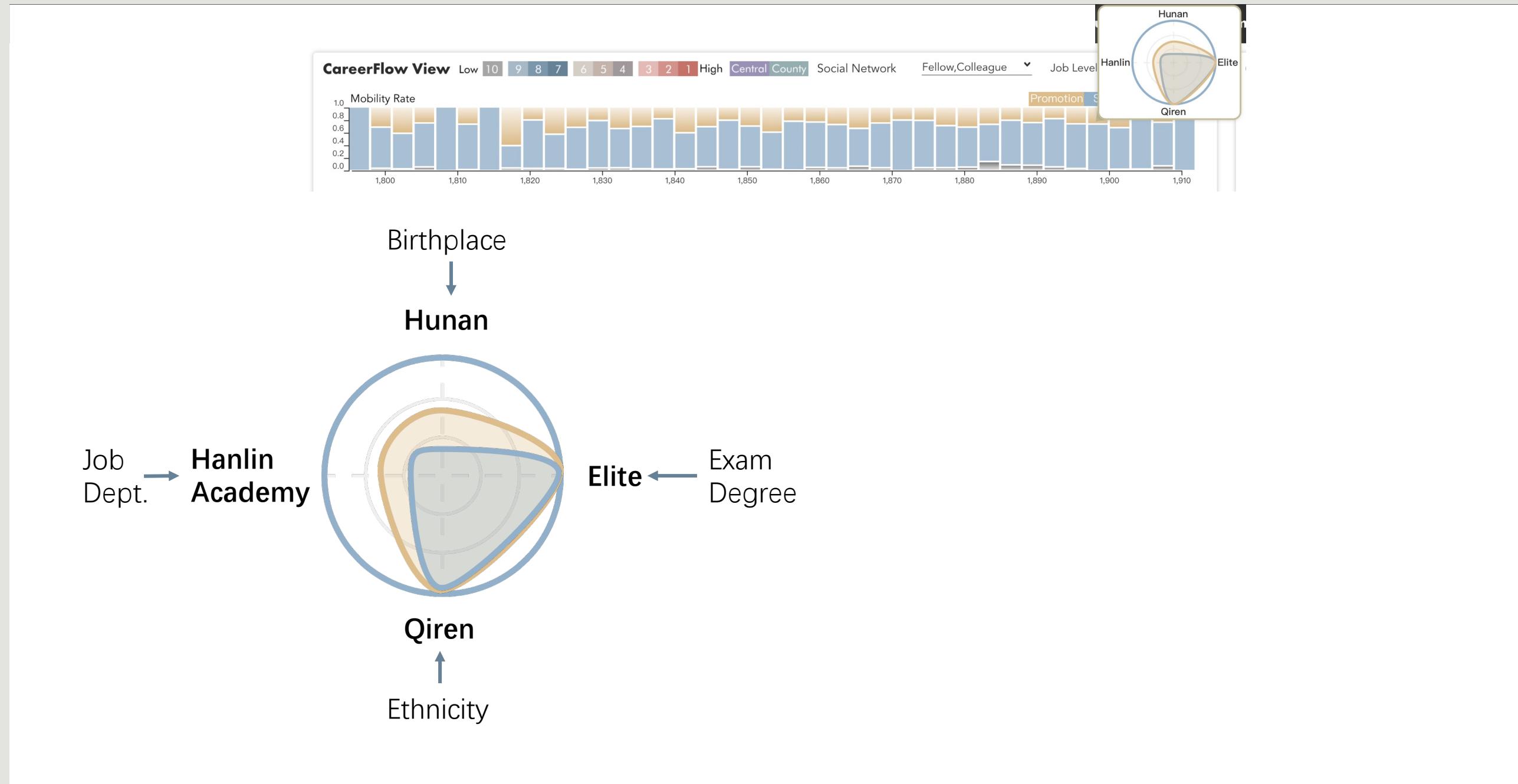
# Visual Design



# Visual Design



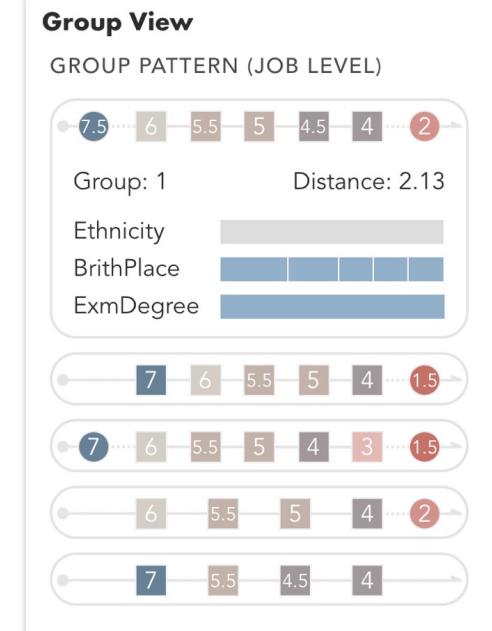
# Visual Design



# Visual Design

## Latent Group Detection (MinDL<sup>1</sup>)

$$L(C) = \sum_{(P,G) \in C} \|P\| + \left( \alpha \sum_{(P,G) \in C} \sum_{s \in G} \|edits(s, P)\| \right) + \lambda \|C\|$$

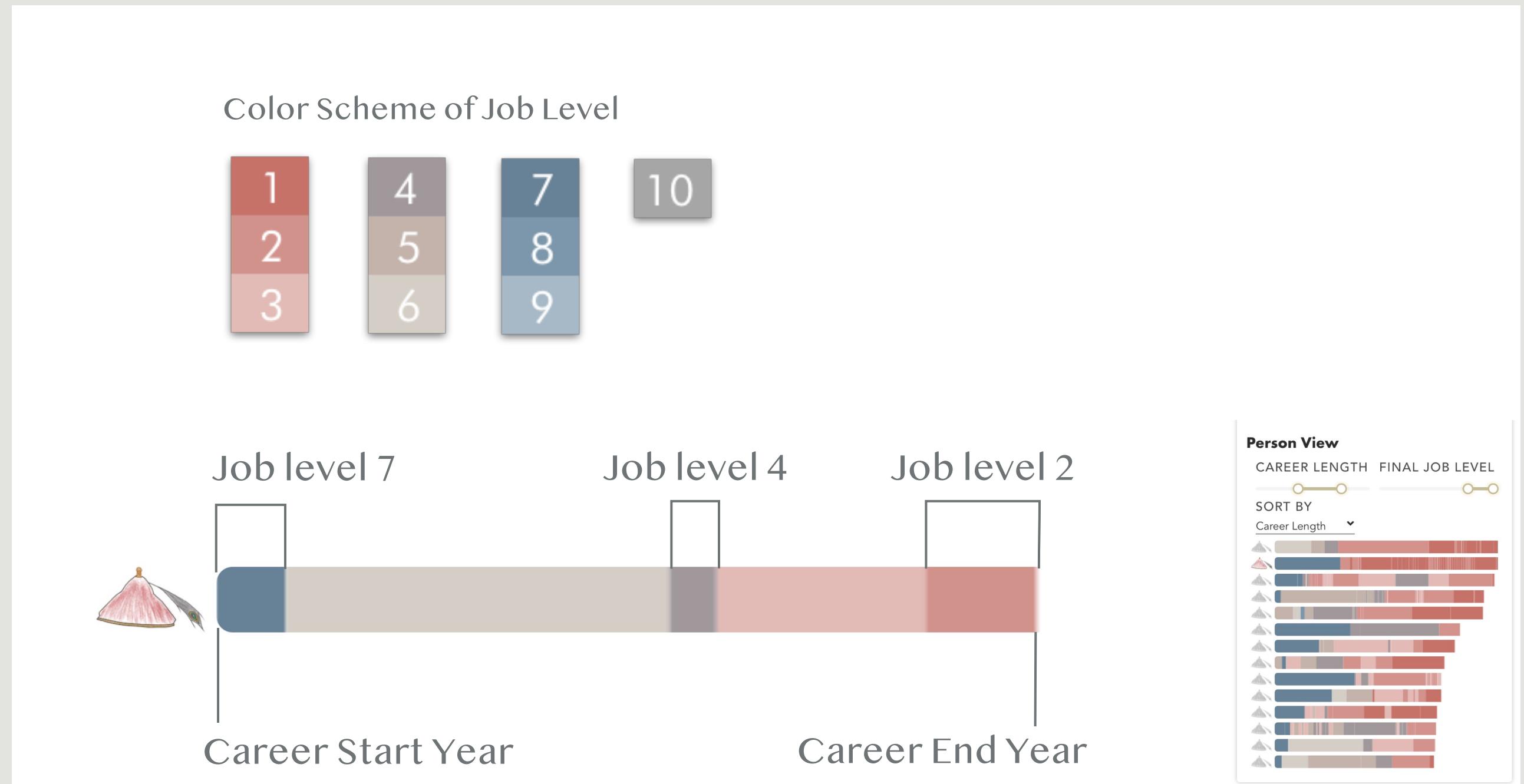


Lowest Job Level    Group Pattern    Highest Job Level

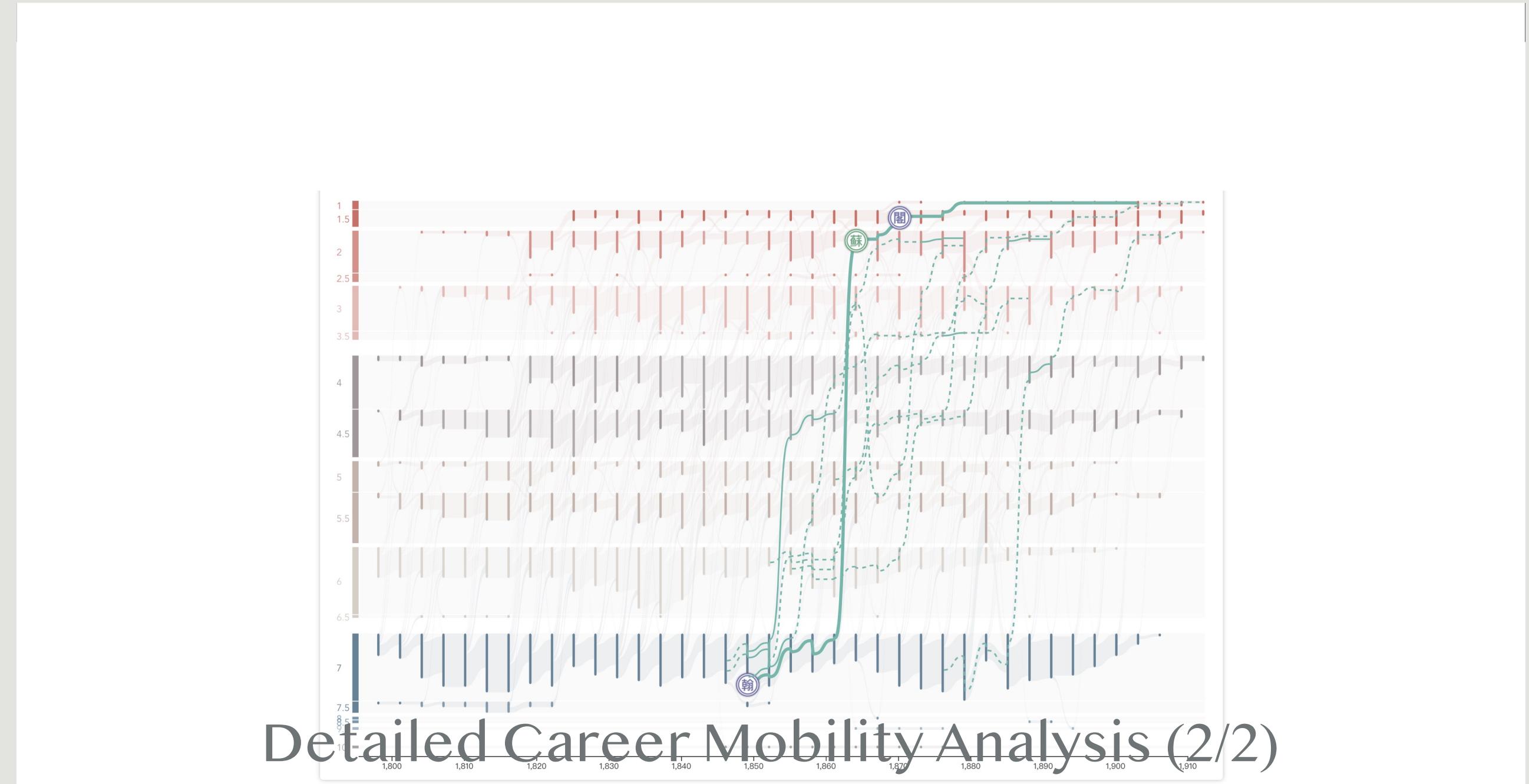


Proportions  
of  
Attribute Values

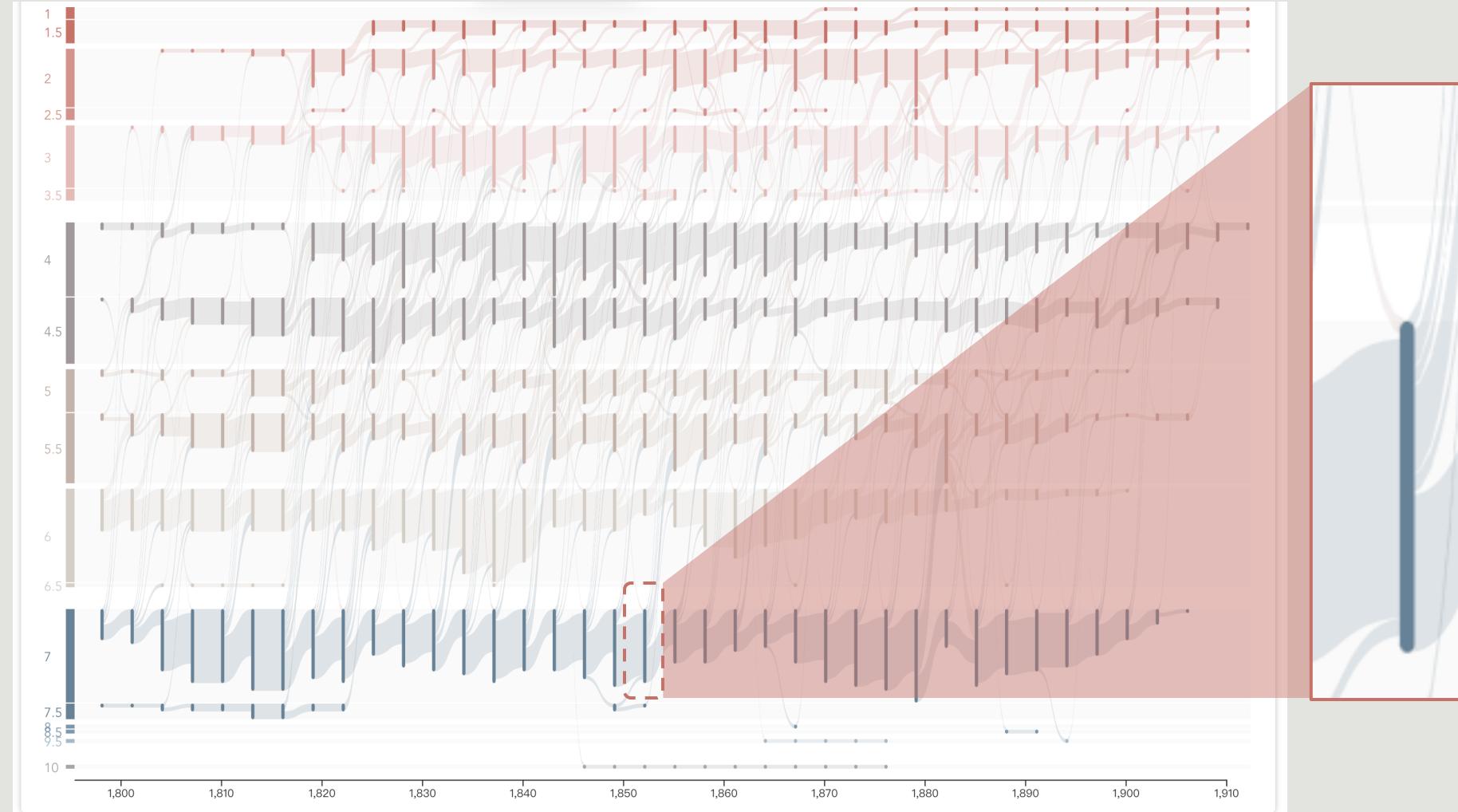
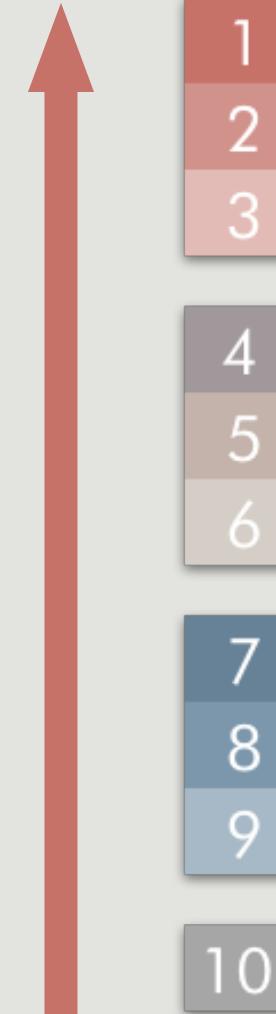
# Visual Design



# Visual Design



# Visual Design



Job-level Mode

# Visual Design

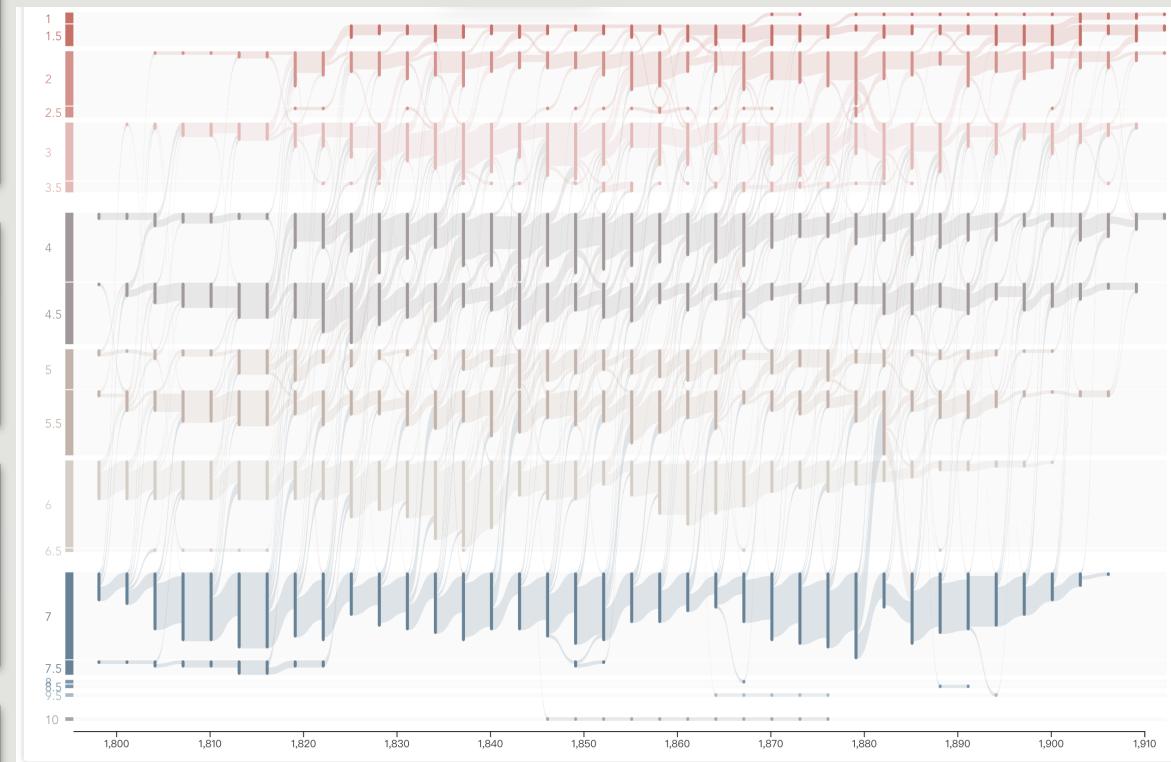


1  
2  
3

4  
5  
6

7  
8  
9

10



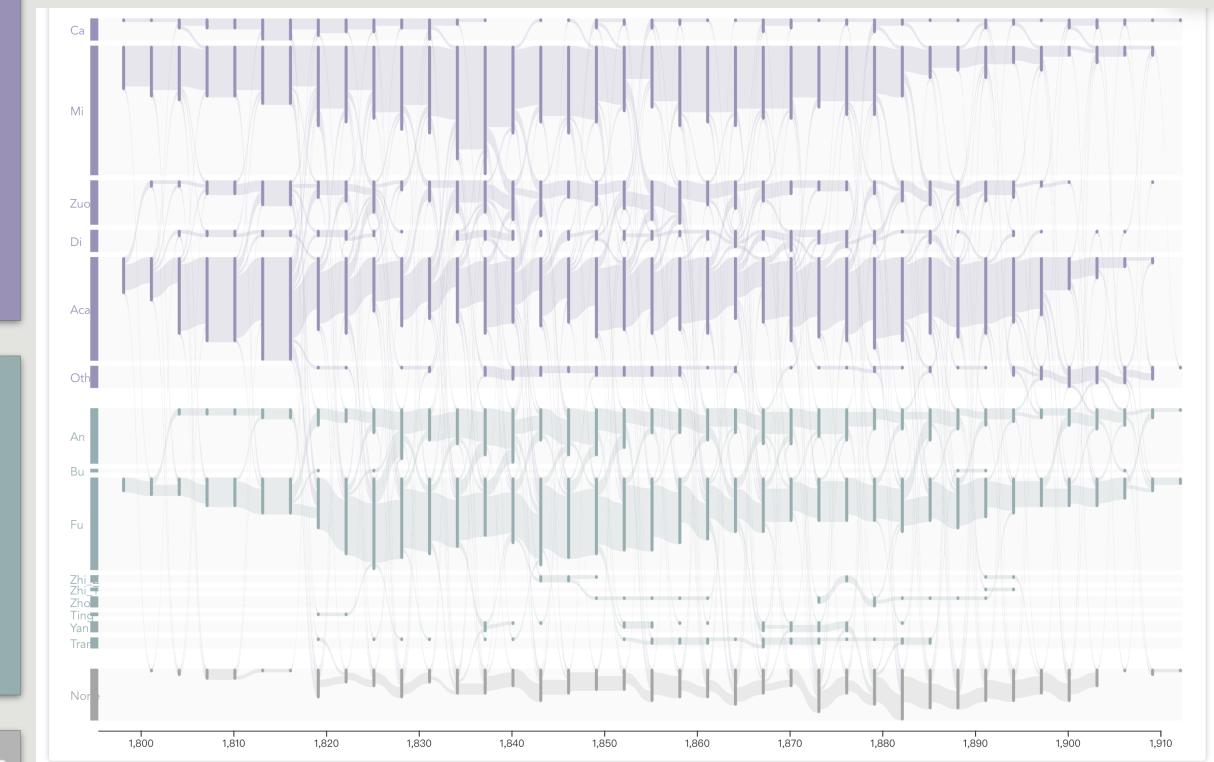
Job-level Mode



Central Gov

County Dept.

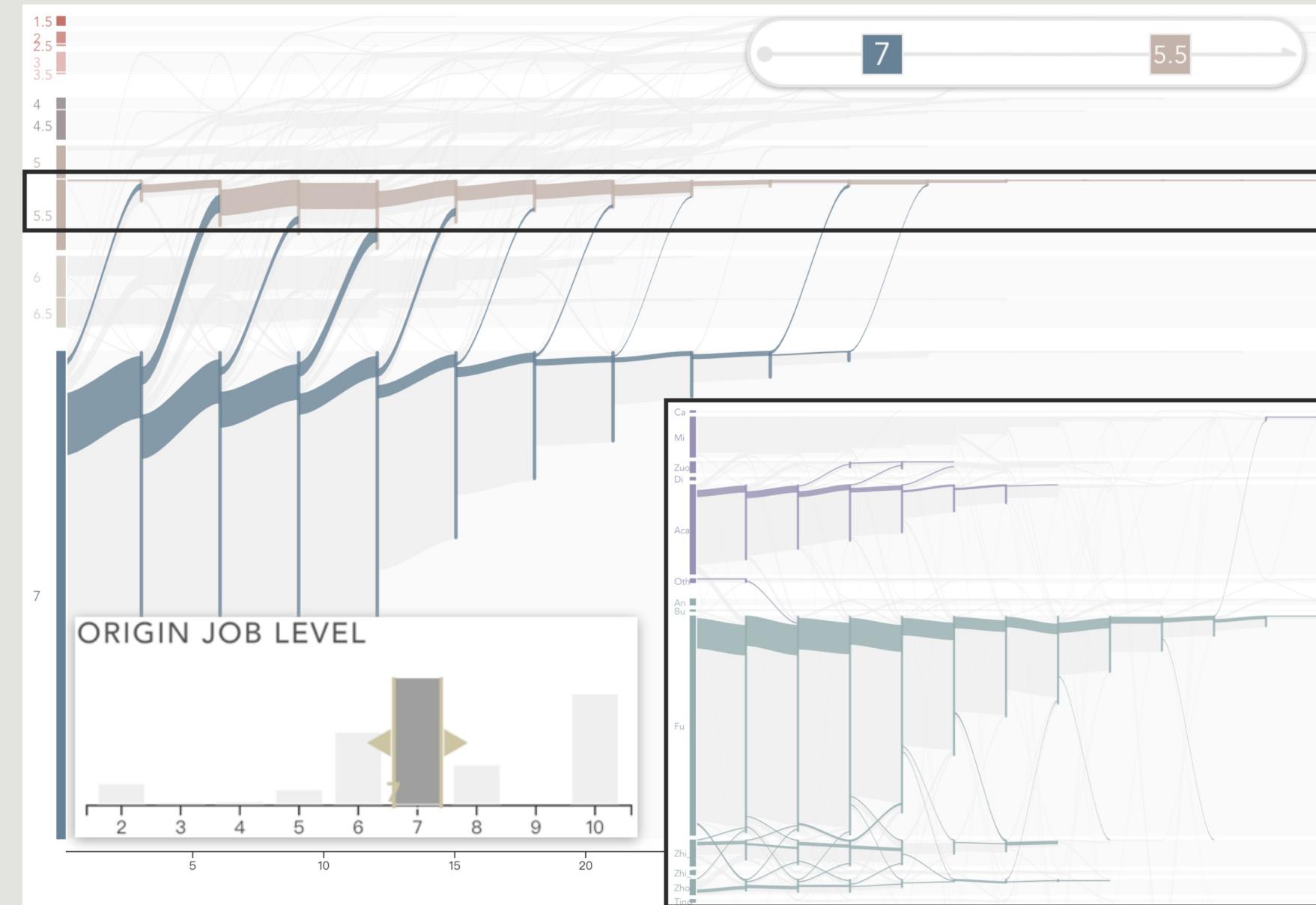
None



Department Mode

- Color Scheme

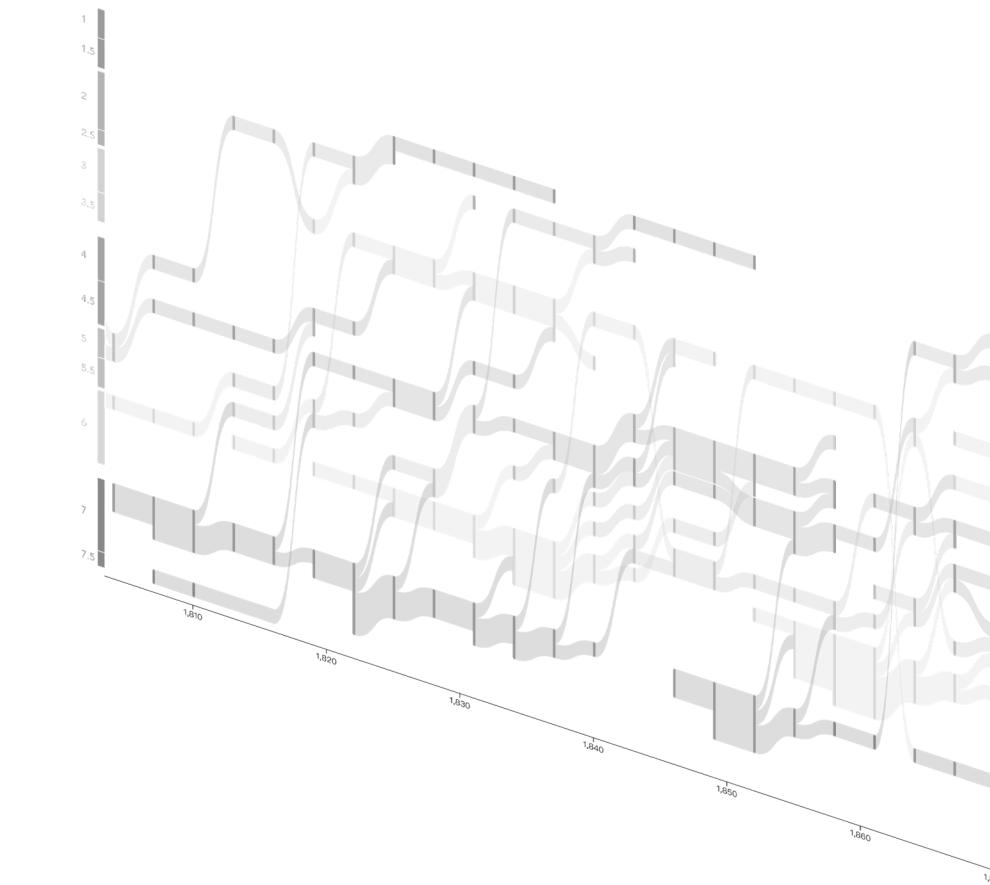
# Visual Design



Relative-time Mode: align the career starting year

# Visual Design

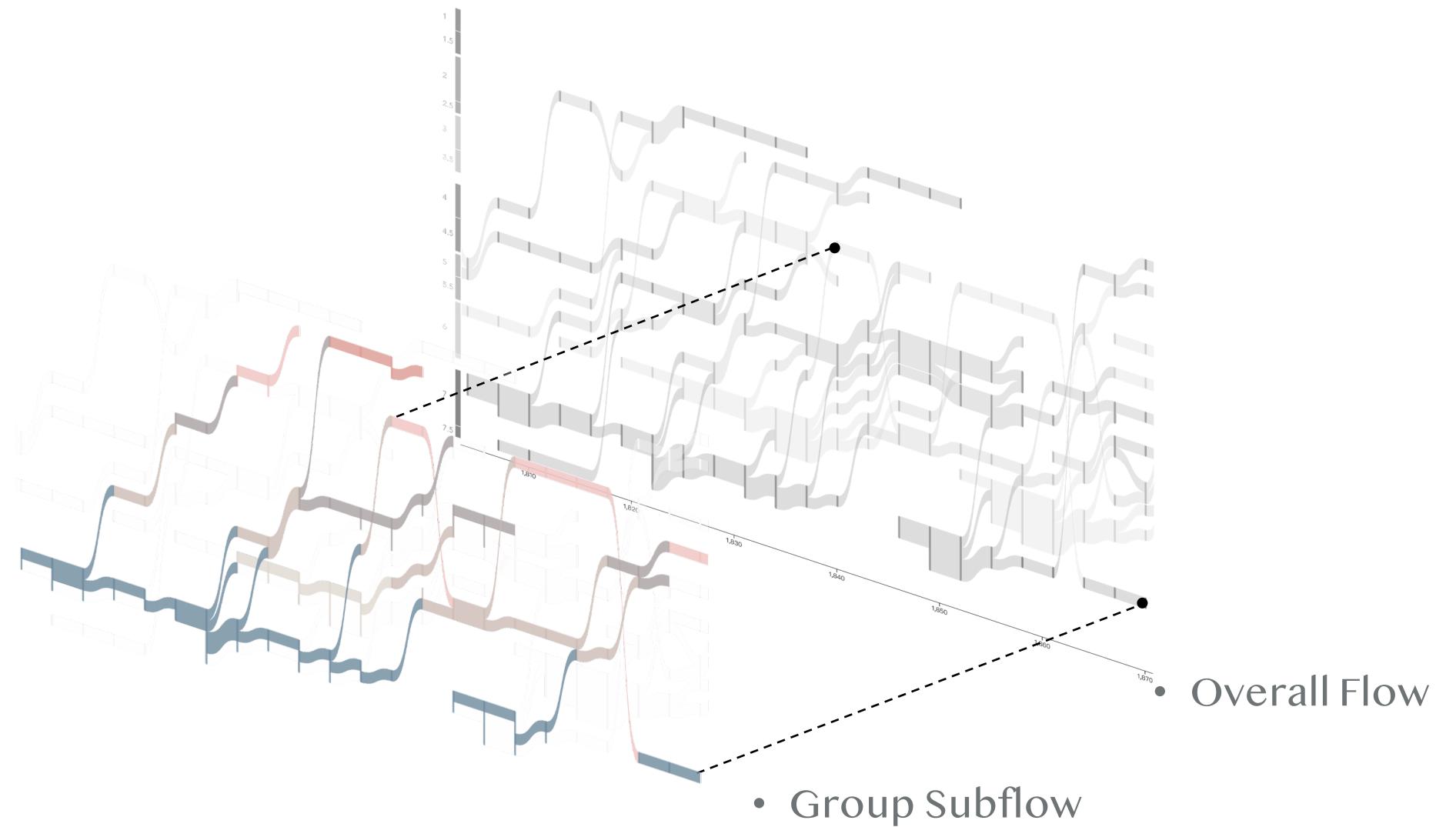
- Flow Design



- Overall Flow

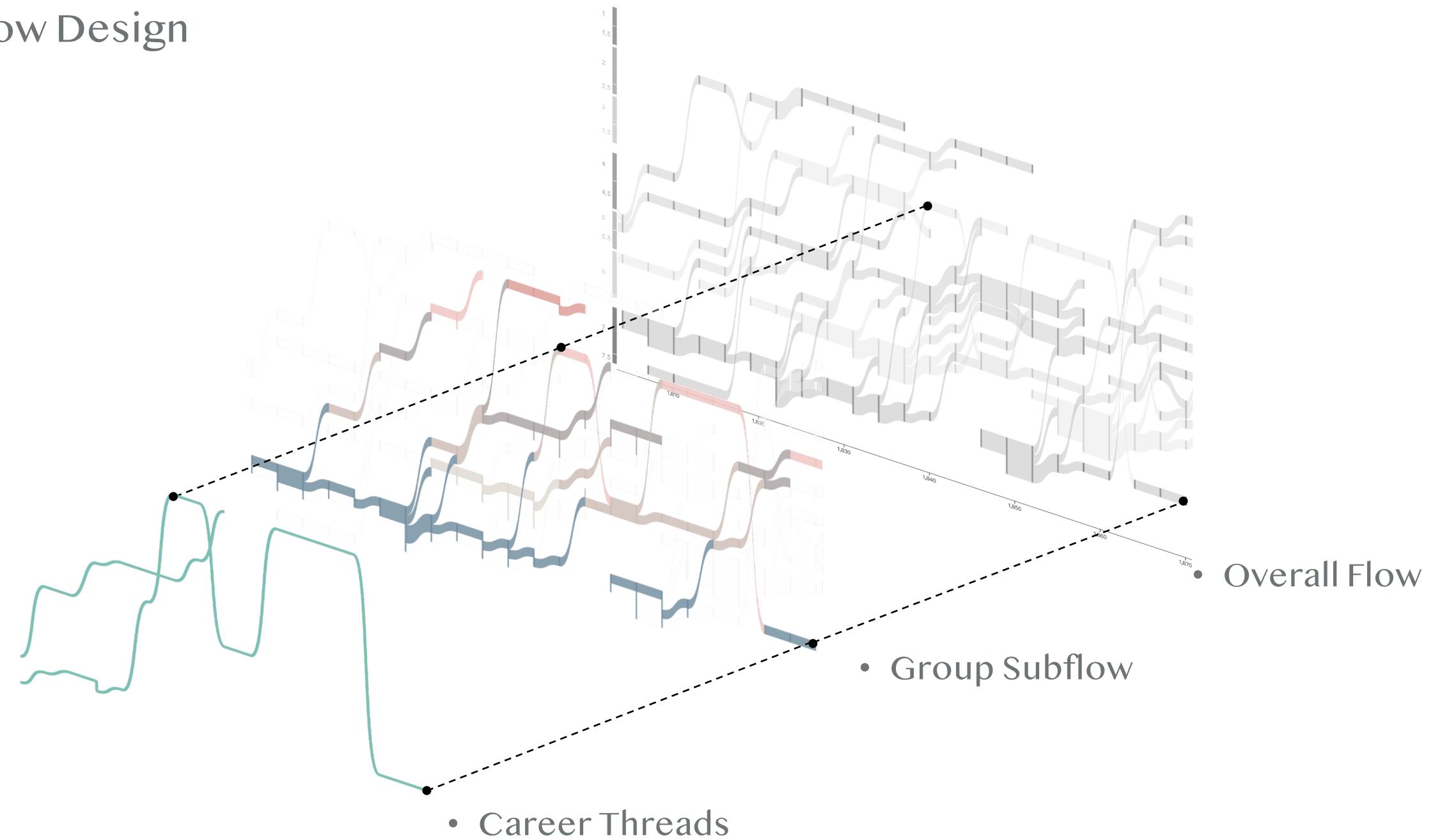
# Visual Design

- Flow Design



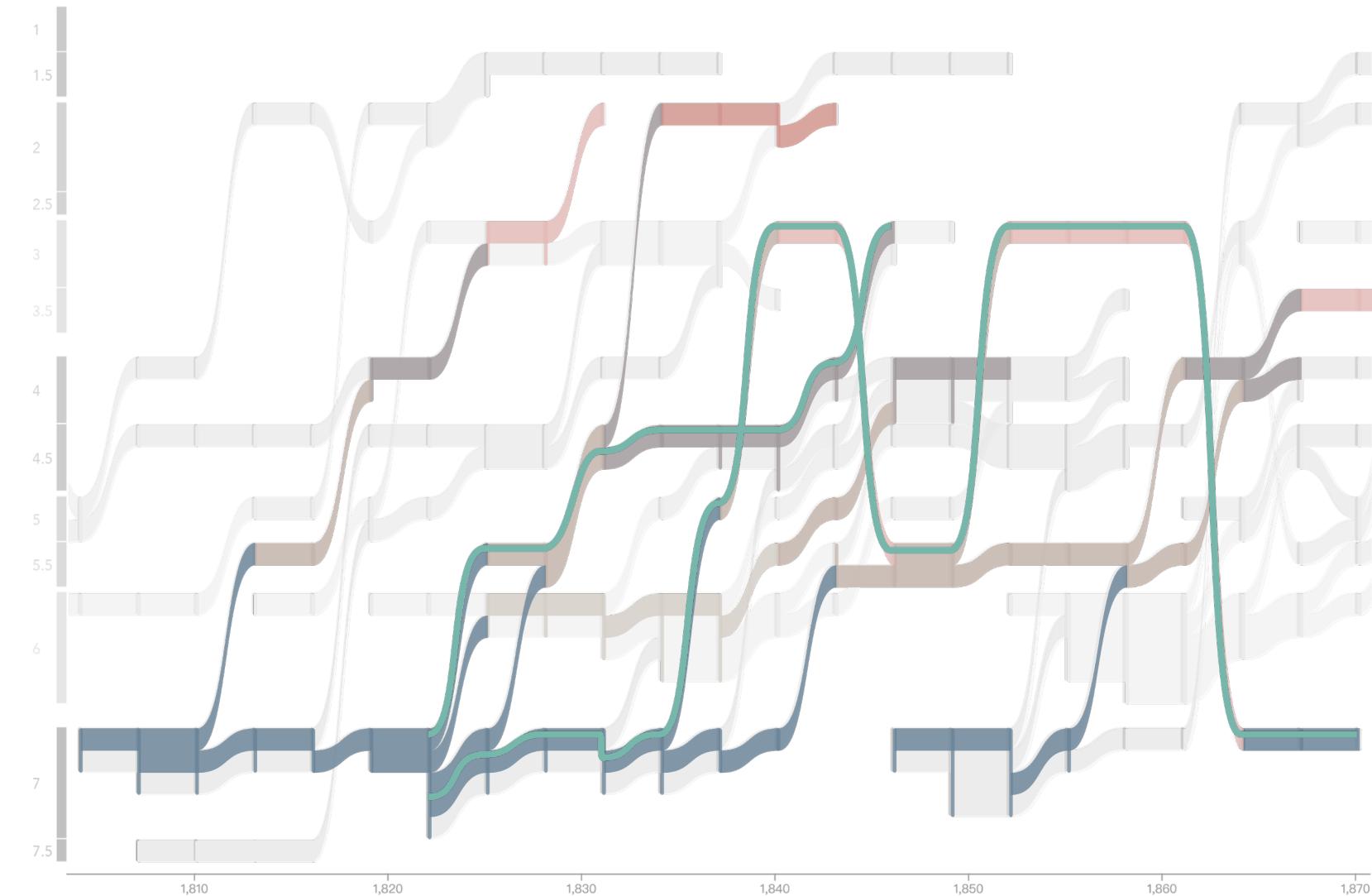
# Visual Design

- Flow Design

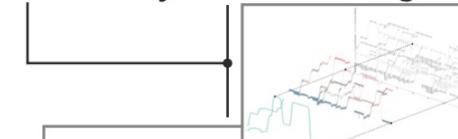


# Visual Design

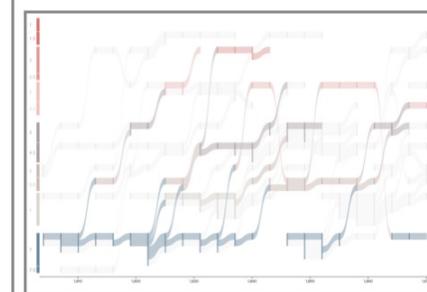
- Flow Design



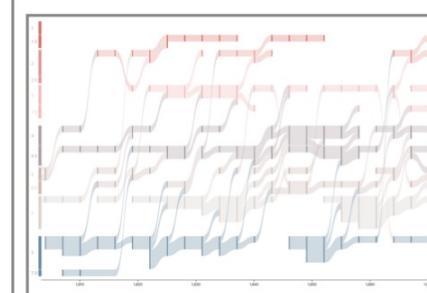
*Three-layer Flow Design*



**Career Thread**  
Highlight several particular individuals



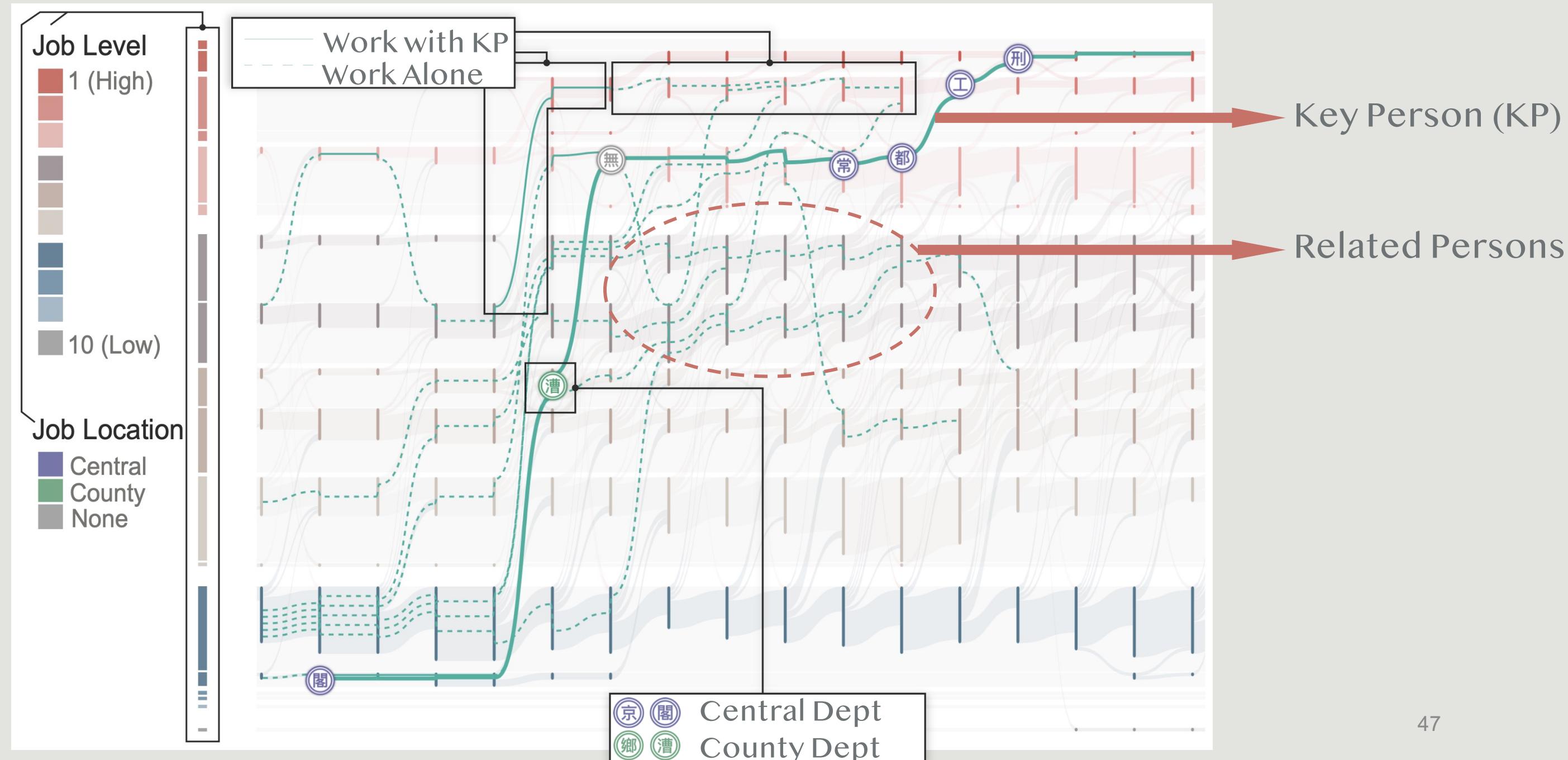
**Group Subflow**  
Aggregate a group into a subflow



**Mobility Flow**  
Provide a context as overall mobility

# Visual Design

- Flow Design + Social Relations of the Key Person (KP)



Case Study  
Expert Interview  
Longitudinal Study

# Evaluation



# Case Study

- Invite four internal experts to freely explore the system
- Political Elites (i.e., 进士) from South and East China

In this video, we demonstrate a case study to explore the political elites from South and East China.

Conclusion

Future Work

Acknowledgement

Q&A



# Conclusion

# Conclusions

## 💡 Contributions

- A hierarchical problem characterization of historical career mobility analysis
- A visual analytics system *CareerLens*
- A novel flow design with a multi-scale approach (overall mobility flow, group subflow, and individual career threads)
- Two case studies and a longitudinal investigation to evaluate the system and receive positive feedback

## 📝 Future Works

- Open to the public
- Multi-attributes for latent group detection
- Improvement of scalability of flow design
- Correlation of career mobility and different events

# For More Information

- I'm currently looking for collaborators in both humanity and social science to work together! For more information, please visit: [wangyifang.top](http://wangyifang.top) or contact [yifang.wang@connect.ust.hk](mailto:yifang.wang@connect.ust.hk)
- The work is partially supported by Hong Kong Research Grants Council (RGC) General Research Fund (GRF) grant 16213317, National Natural Science Foundation of China (62072400), Zhejiang Provincial Natural Science Foundation (LR18F020001), and the 100 Talents Program of Zhejiang University.  
Construction of the CGED-Q was supported by Hong Kong RGC GRF 16600017



# Interactive Visual Exploration of Longitudinal Historical Career Mobility Data



Yifang Wang<sup>1,2</sup>



Hongye Liang<sup>2,3</sup>



Xinhuan Shu<sup>1</sup>



Jiachen Wang<sup>2,3</sup>



Ke Xu<sup>4</sup>



Zikun Deng<sup>2,3</sup>



Cameron Campbell<sup>1</sup>



Bijia Chen<sup>5</sup>



Yingcai Wu<sup>2,3</sup>



Huamin Qu<sup>1</sup>

<sup>1</sup> 香港科技大学  
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY



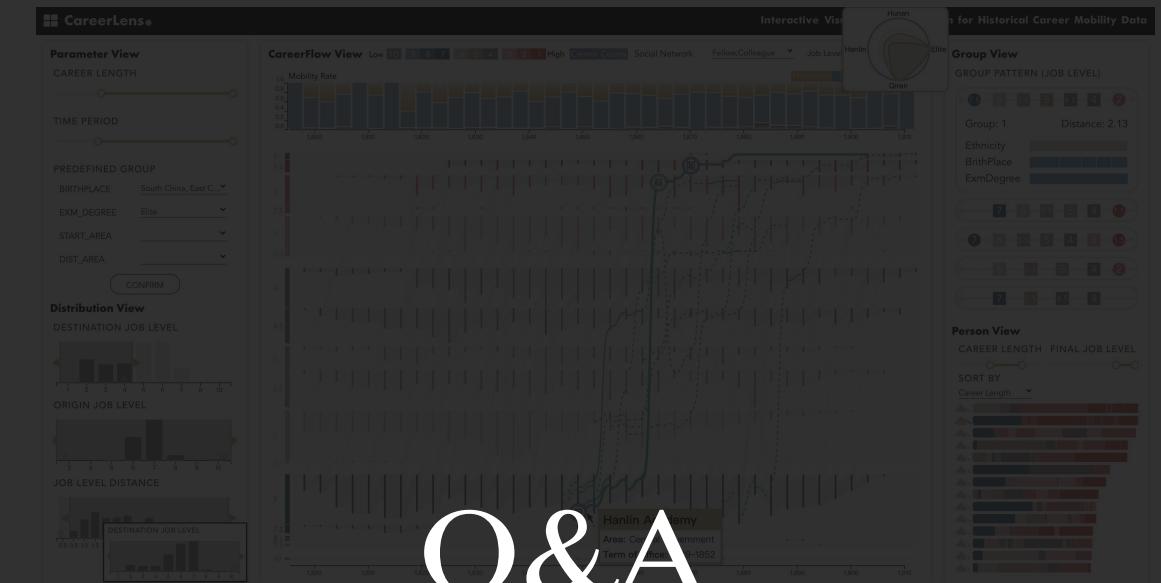
<sup>2</sup> 浙江大学  
ZHEJIANG UNIVERSITY

<sup>3</sup> 之江实验室  
ZHEJIANG LAB

<sup>4</sup> NYU



<sup>5</sup> 中国农业大学  
RENMIN UNIVERSITY OF CHINA



Q&A

## Data Preprocessing



- I'm currently looking for collaborators in both social science and humanities to work together!



- For more information, please visit: [wangyifang.top](http://wangyifang.top) or contact [yifang.wang@connect.ust.hk](mailto:yifang.wang@connect.ust.hk)