



School of Economics and Management

Information Science III

1. Introduction

Yuki Yanai



<https://yukiyanai.github.io>



yanai.yuki@kochi-tech.ac.jp



Today's Goals

- To understand:
 - Outline of the course
 - Why data visualization matters

Course Outline

Class format

- Lecture
- Discussion
- Computer lab
 - Visualization exercises
 - R Programming (mainly using ggplot2)

Textbooks

- Healy, K. 2019. *Data Visualization: A Practical Introduction*. Princeton UP. ([Draft Version](#))
- Kirk, A. *Data Visualisation: A Handbook for Data Driven Design*, 2nd Edition. SAGE.
- Wilke, C. O. 2019. [Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures](#). O'Reilly.
- Wickham, H. 2016. [ggplot2: Elegant Graphics for Data Analysis, 2nd Edition](#). Springer.
- Yau, N. 2011. *Visualize This: The FlowingData Guide to Design, Visualization, and Statistics*. Wiley.

Reference Books

- Please refer to the official online syllabus
- Required readings will be distributed via KUTLMS (moodle)

Grade

- Grades will be based on:
 - Participation in class activities [20%]
 - (Almost) Weekly assignments [40%]
 - In-class presentation [10%]
 - Final project [30%]
- Please refer to the official online syllabus for grading criteria

R

- I assume that you know how to use R
 - ▶ Prerequisites: Statistics 2 (統計学2) *and* Programming (プログラミング)
- This course doesn't teach you the basics of R
 - ▶ Please learn it by yourself
 - ▶ Or you cannot:
 - participate fully in class activities
 - Complete weekly assignments

Office hours

- Time: 4:50 - 6:20 pm on Tuesdays
- Place: A625
- You may talk in Japanese during house hours
- Please make an appointment in advance if you'd like to visit my office other than the office hours

KUTLMS (Moodle)

- URL: <https://lms.kochi-tech.ac.jp/course/view.php?id=1968>
- Registration key: Hadley2022

Why Visualization?

Better Communication

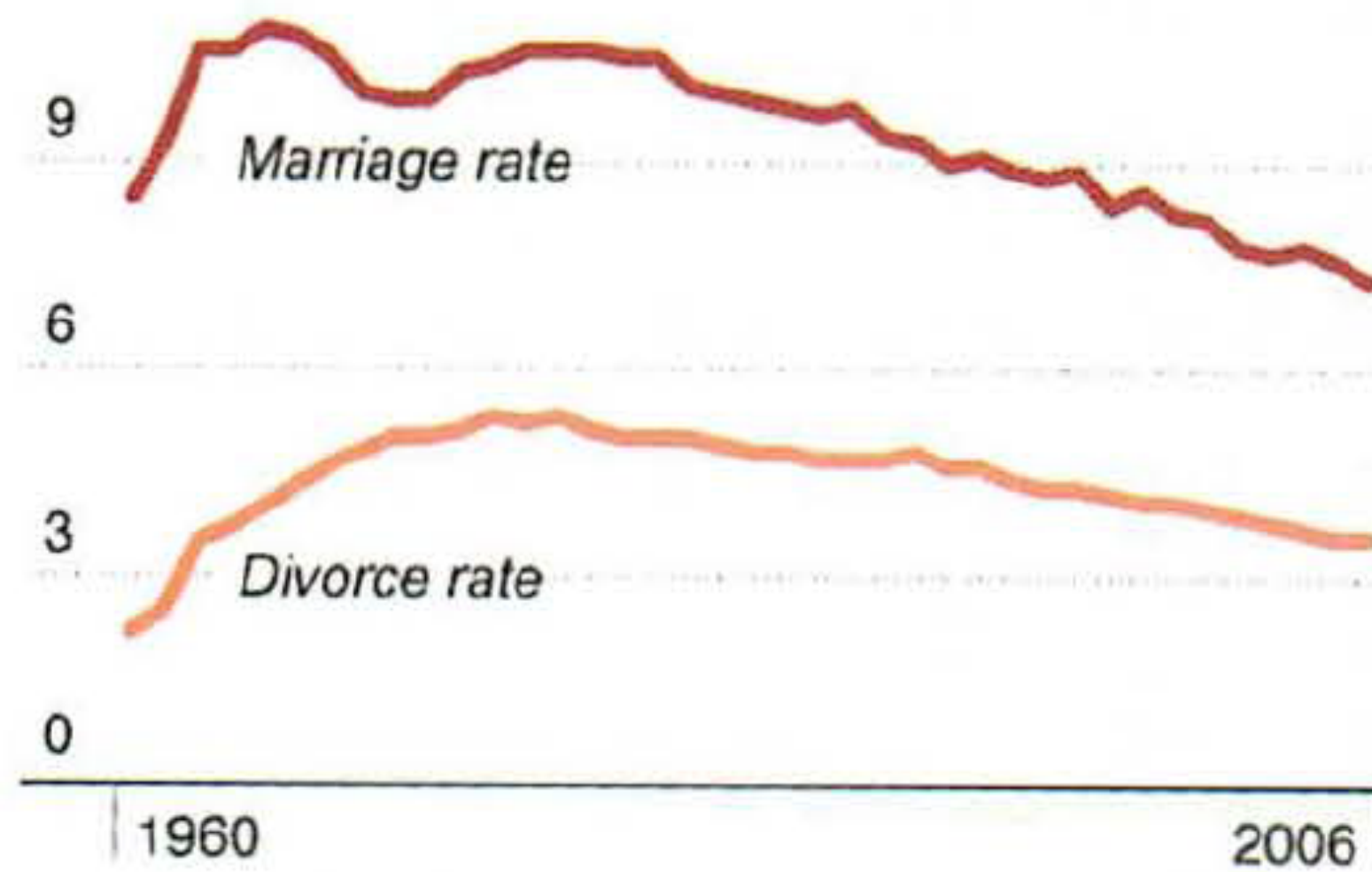
- Visualization helps us:
 - Understand data better
 - Clarify patterns in data
- More convincing
 - Good visualization makes it easier to communicate with audience
 - Bad visualization:
 - Misleads or confuses people
 - Could transmit “wrong” ideas

How to Present Data

- Traditional way: Tables
 - Accurate
 - Easy to make (?)
 - Most people know how to read “tables”
- But...
- Let's discuss how and what we should learn from a table (handout)

Marriage and Divorce, 1960-2006

12 per 1,000 population



Yau (2011, p. XX)

Shift from Tables to Figures

- Kastellec, J. P., and E. L. Leoni. 2007. “Using Graphs Instead of Tables in Political Science.” *Perspective on Politics* 5(4): 755-711 [[PDF](#)]
- Some R Codes by Fredrick Solt: https://fsolt.org/dotwhisker/articles/kl2007_examples.html

Example: Table

Table 1

**Iversen and Soskice 2006, table 1:
Electoral system and the number of years
with left and right governments (1945–98)**

		Government Partisanship		Proportion of Right Governments
		Left	Right	
Electoral system	Proportional	342 (8)	120 (1)	.26
	Majoritarian	86 (0)	256 (8)	.75

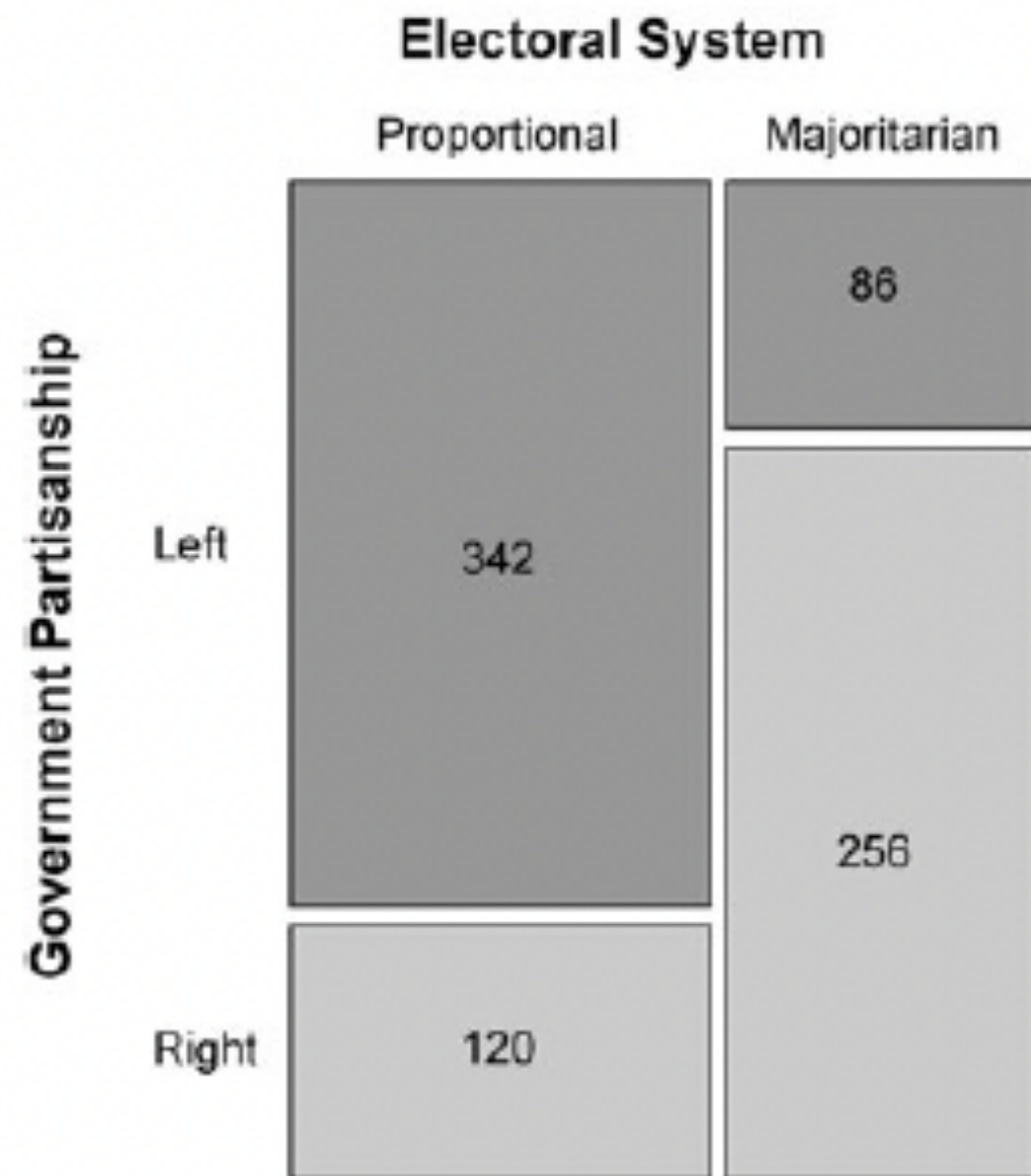
Example: Mosaic Plot

Table 1

Iversen and Soskice 2006, table 1:
Electoral system and the number of years
with left and right governments (1945–98)

		Government Partisanship		
		Left	Right	Proportion of Right Governments
Electoral system	Proportional	342 (8)	120 (1)	.26
	Majoritarian	86 (0)	256 (8)	.75

**Electoral Systems and
Government Partisanship,
1945–1998**



Example: Table

Table 2

McClurg 2006, table 1 (panel A): The political character of social networks

	Mean	Standard Deviation	Min	Max	N
Panel A: Descriptive Statistics					
Size ^a	3.13	1.49	1	5	1260
Political Talk	1.82	0.61	0	3	1253
Political Agreement	0.43	0.41	0	1	1154
Political Knowledge	1.22	0.42	0	2	1220

Notes: This table provides descriptive statistics for the political character of the social networks as perceived by respondents.

^aWhen respondents who report having *no network* are included the mean of this variable drops to 2.57 with a standard deviation 1.81 (n = 1537).

Example: Dot-and-whisker plot

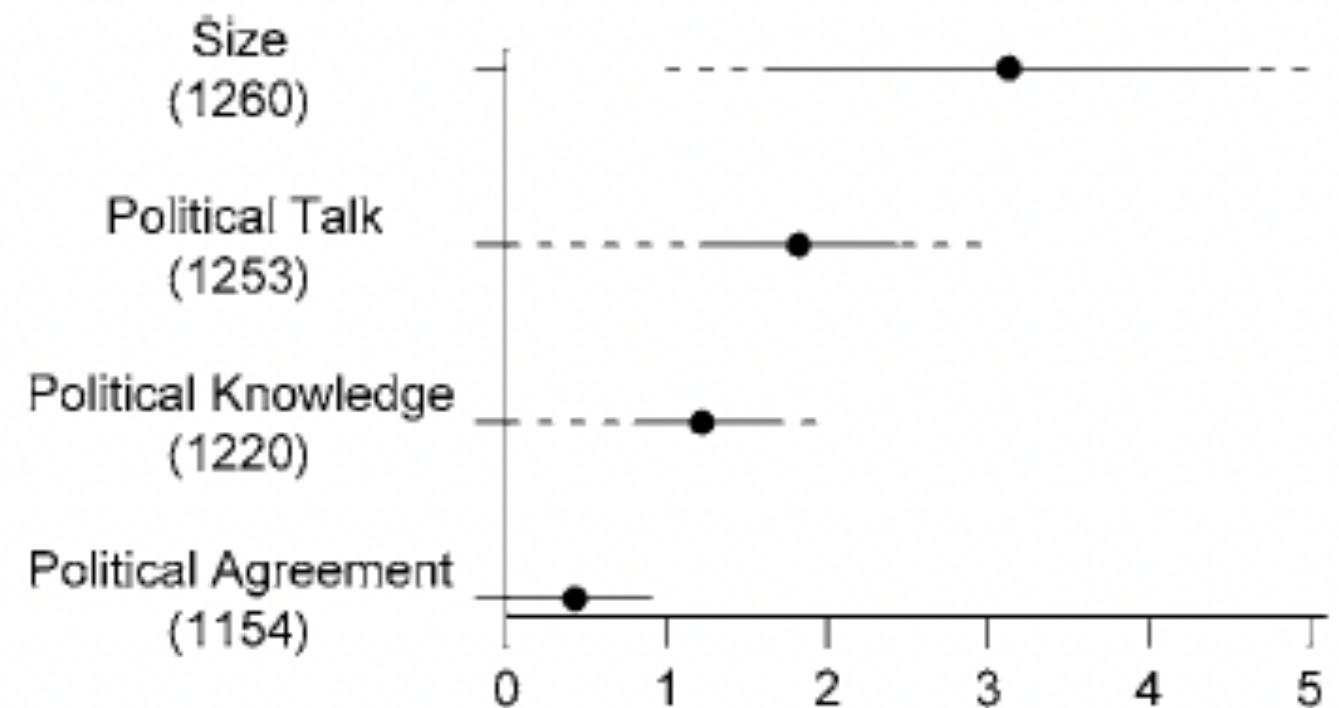
Table 2
McClurg 2006, table 1 (panel A): The political character of social networks

	Mean	Standard Deviation	Min	Max	N
Panel A: Descriptive Statistics					
Size ^a	3.13	1.49	1	5	1260
Political Talk	1.82	0.61	0	3	1253
Political Agreement	0.43	0.41	0	1	1154
Political Knowledge	1.22	0.42	0	2	1220

Notes: This table provides descriptive statistics for the political character of the social networks as perceived by respondents.

^aWhen respondents who report having *no network* are included the mean of this variable drops to 2.57 with a standard deviation 1.81 (n = 1537).

Figure 3
Using a Single Dot Plot to Present Summary Statistics.



Example: Table (Regression Results)

Table 8
Pekkanen, Nyblade and Krauss (2006),
table 1: Logit analysis of electoral
incentives and LDP post allocation
(1996–2003)

Variable	Model 1	Model 2
<i>Block 1: MP Type</i>		
Zombie	0.18 (.22)	0.27 (0.22)
SMD Only	−0.19 (0.22)	−0.19 (0.24)
PR Only	−0.39 (0.18)**	—
Costa Rican in PR	−0.09 (0.29)	—
<i>Block 2: Electoral Strength</i>		
Vote share margin	—	0.005 (0.004)
Margin Squared	—	—
<i>Block 3: Misc Controls</i>		
Urban-Rural Index	0.04 (0.08)	0.04 (0.09)
No Factional	−0.86 (0.26)***	−0.98 (0.31)***
Membership		
Legal Professional	0.39 (0.29)	−.36 (0.30)
<i>Seniority</i>		
1 st Term	−3.76 (0.36)***	−3.66 (0.37)***
2 nd Term	−1.61 (0.19)***	−1.59 (0.21)***
4 th Term	−0.34 (0.19)**	−0.45 (0.21)***
5 th Term	−1.17 (0.22)***	−1.24 (0.24)***
6 th Term	−1.15 (0.22)***	−1.04 (0.24)***
7 th Term	−1.52 (0.25)***	−1.83 (0.29)***
8 th Term	−1.66 (0.28)***	−1.82 (0.32)***
9 th Term	−1.34 (0.32)***	−1.21 (0.33)***
10 th Term	−2.89 (0.48)***	−2.77 (0.49)***
11 th Term	−1.88 (0.43)***	−1.34 (0.46)***
12 th Term	−1.08 (0.41)***	−0.94 (0.49)**
Constant	.020 (.20)	0.13 (0.26)
Log-likelihood	−917.24	−764.77
<i>N</i>	1895	1574

Notes: Dependent Variables: 1 if MP holds a post of minister, vice minister, PARC, or HoR Committee Chair.

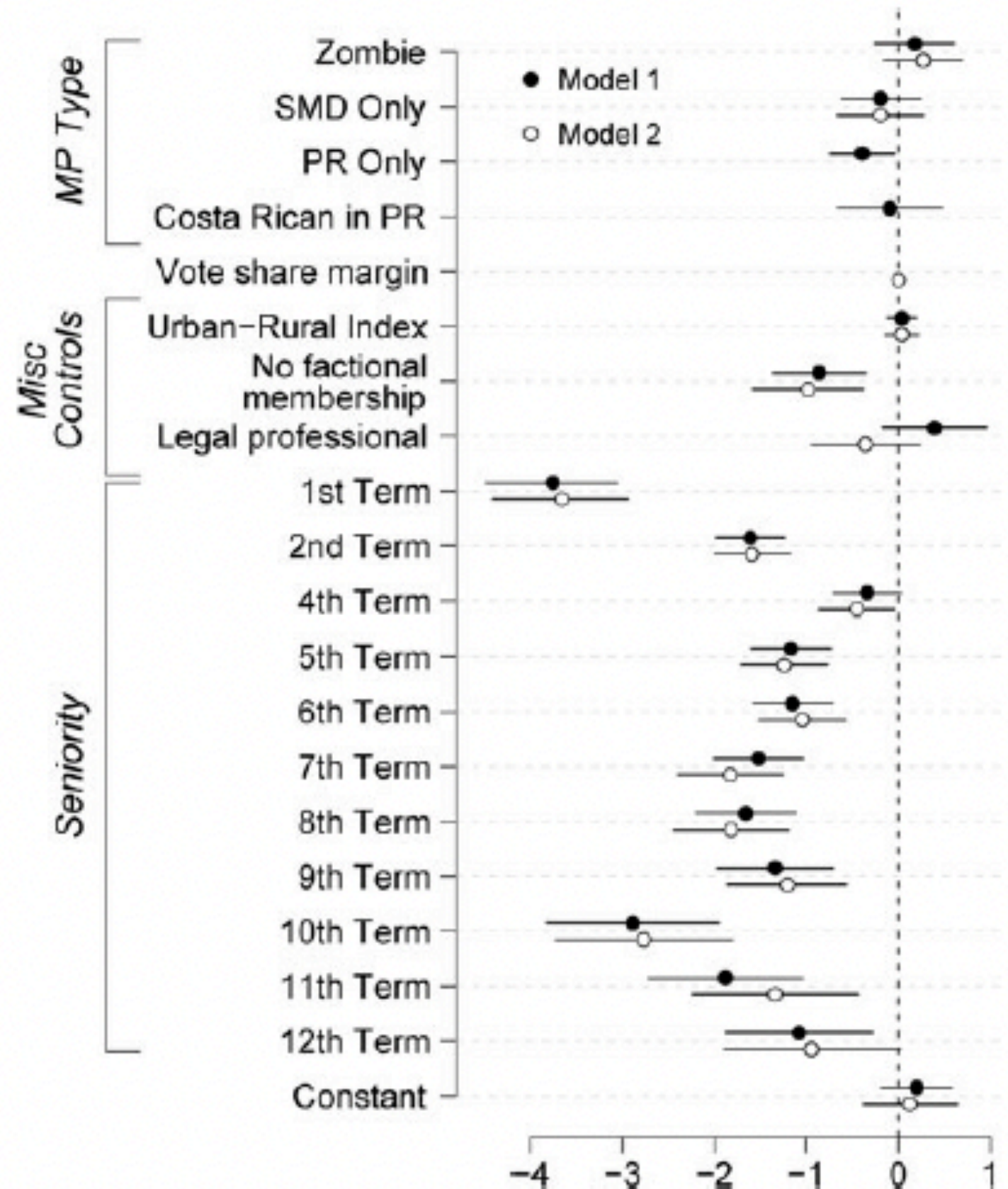
Base categories: SMD dual-listed, 3rd term. Excluded observations: senior MPs that held no post (> 12 terms, PR-Only MPs in Model 2).

*p < .10, **p < .05, ***p < .001.

Example: Dot-and- whisker plot for two models

Figure 7

Using parallel dot plots with error bars to present two regression models.



Dynamic Visualization: An Example

[https://www.ted.com/talks/
hans_rosling_let_my_dataset_change_your_mindset?
language=en](https://www.ted.com/talks/hans_rosling_let_my_dataset_change_your_mindset?language=en)

Visualize Space

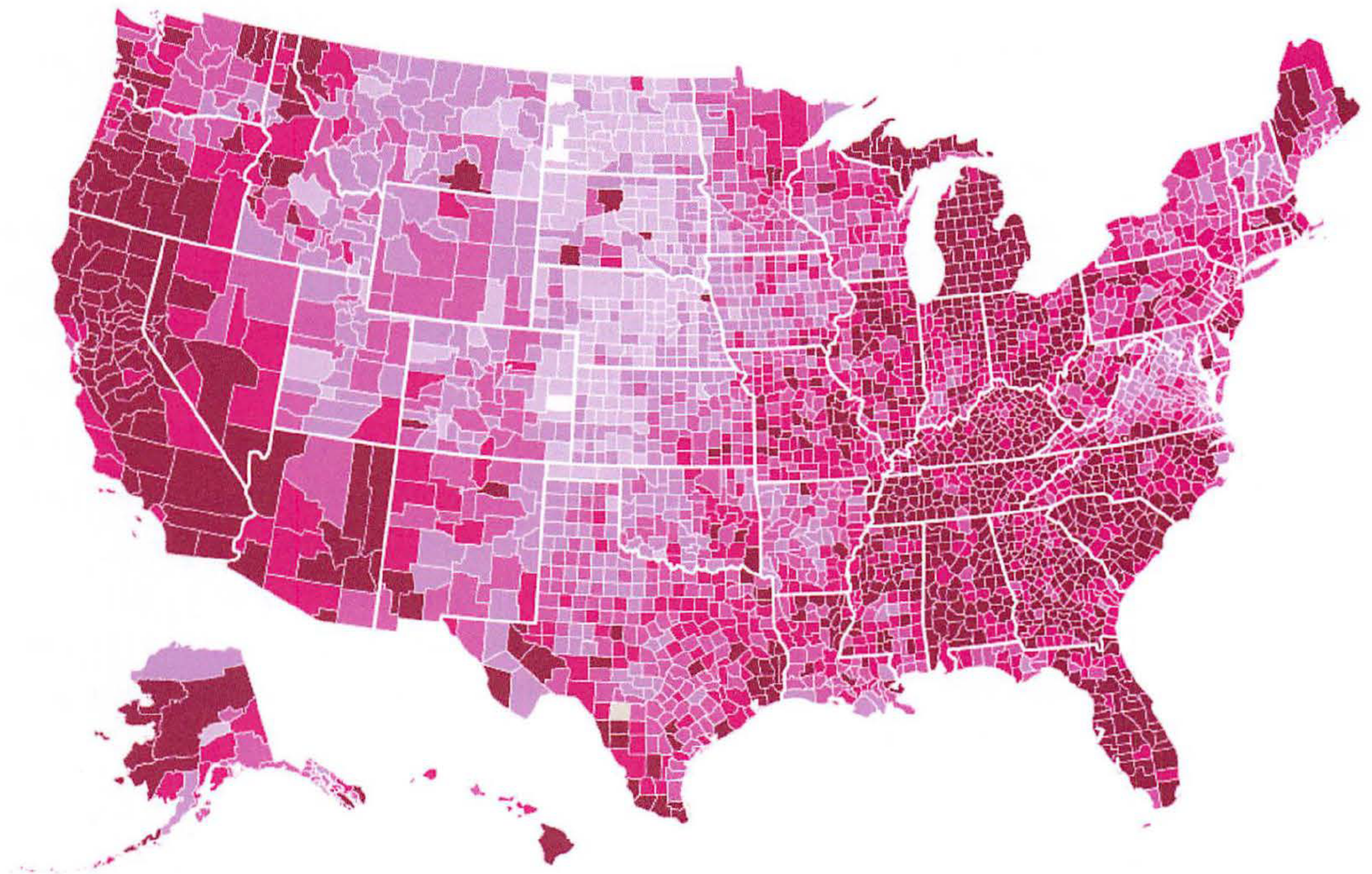


FIGURE I-2 Map of unemployment for 2009

Yau (2011, p. XVII)

Bad Visualization: An Example



Kochi Shimbun, April 24, 2021

You will learn more during the course!

Next class

2. Data