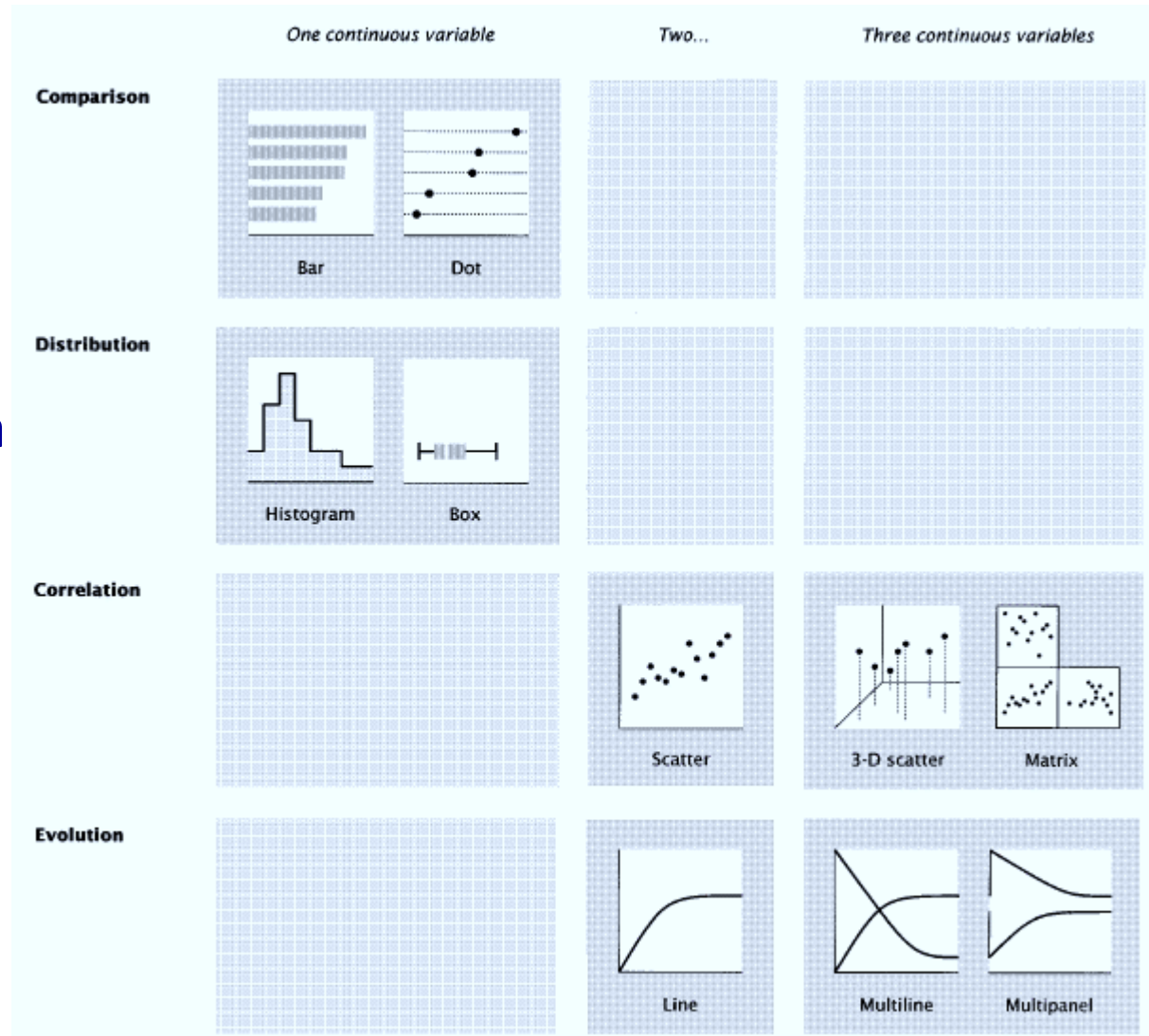


# Relating data structure and graph design

ME447 Visualizing Data  
Spring 2019–20

Richard Layton



# Jean-Luc Doumont: Optimal graph design depends on **the variables** to be shown...



Number of variables?  
Continuous or discrete?



Number of variables?  
Nominal or ordinal?  
Number of levels each?

... and **the message** to be conveyed



**displaying distributions**

**comparing data**

**revealing correlations**

**showing evolution**



# Let's examine match-ups of graph design/data structure

62

strip plot

box and whisker plot

multiway

scatterplot

dot plot

line graph

conditioning plot

scatterplot matrix

63

parallel coordinate plot

cycle plot

mosaic plot

financial (OHLC) plot

diverging stacked bar

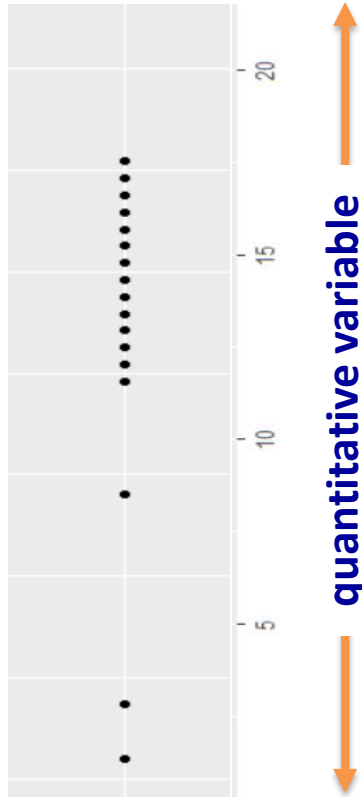
linked micromaps

proportional symbol map

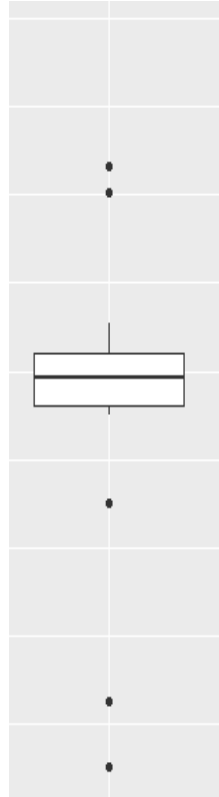
dot density map

Story: **distributions**  
Data: **1 quantitative variable**

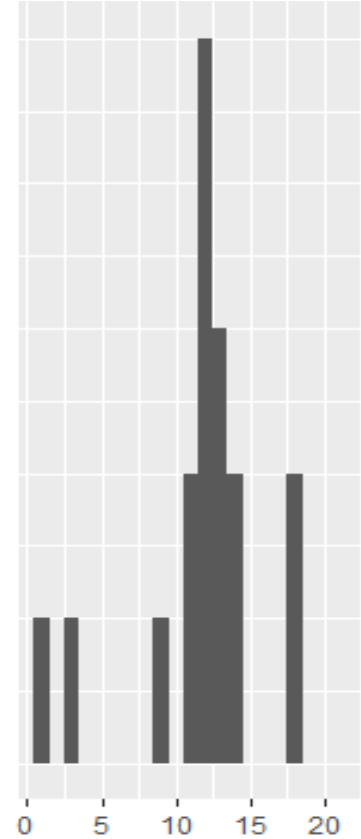
*strip plot or box plot*



**strip plot**  
**preferred**

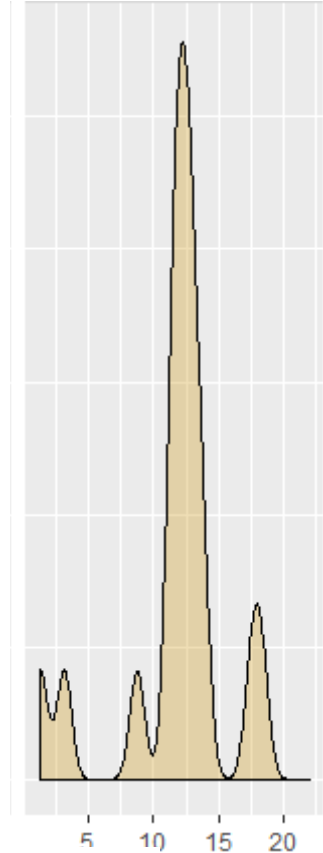


**box & whisker**  
**preferred**



**histogram**

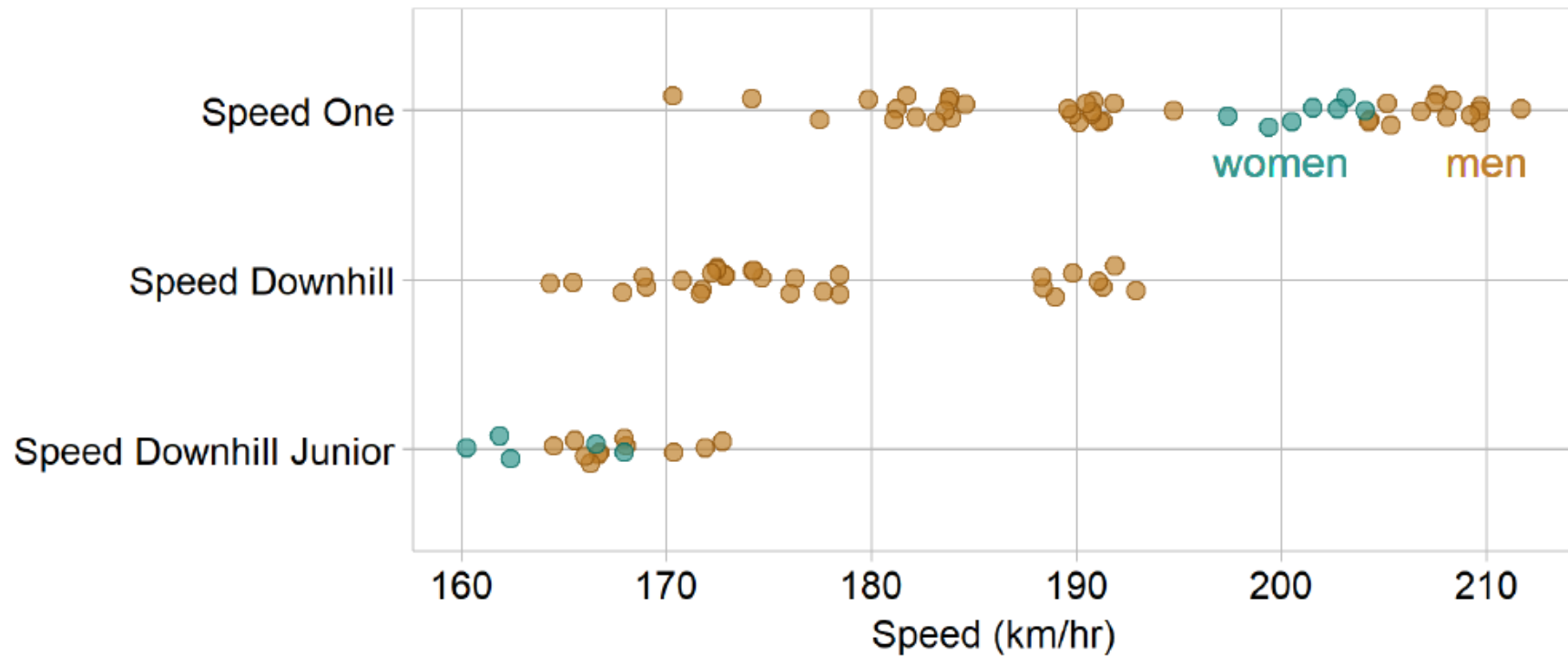
**(prone to rhetorical malpractice)**



**density**

# *strip plot*

Olympic speed skiing



Data source (Unwin, 2015)

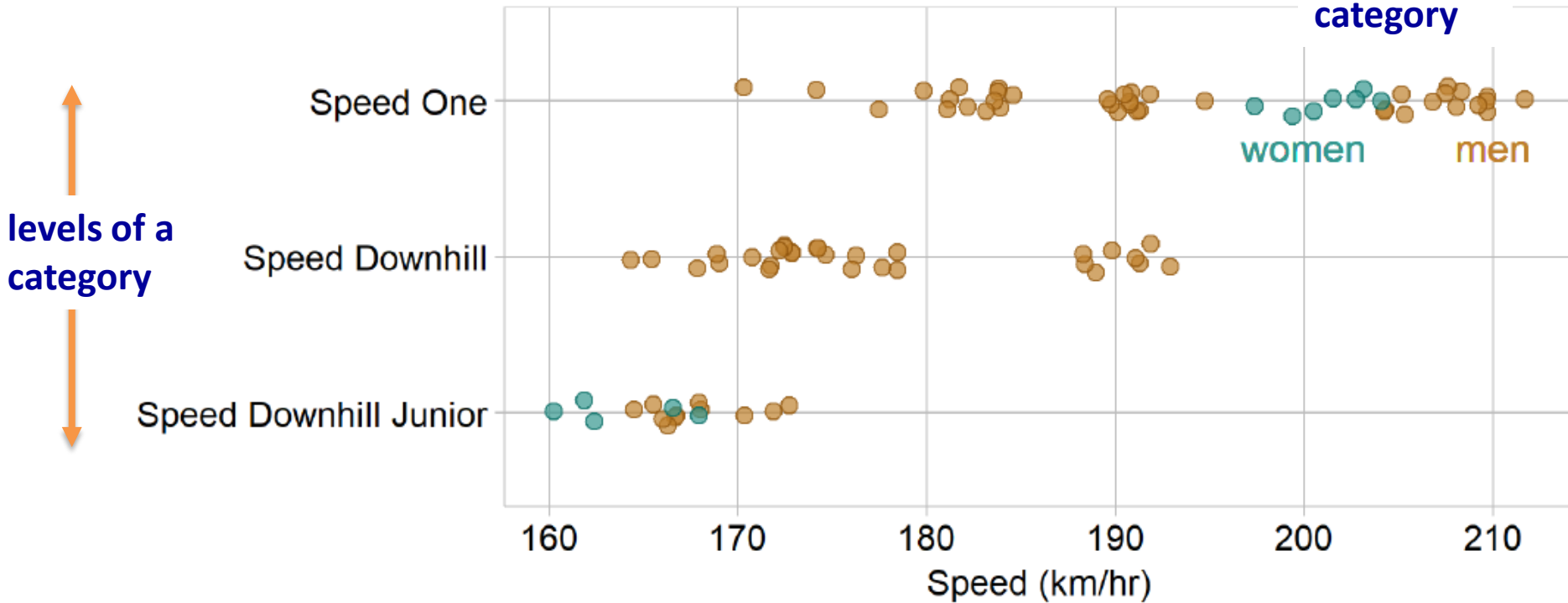
**Story:** distribution and comparison

**Data:** 1 quantitative, 2 categorical

*strip plot*

Olympic speed skiing

levels of a  
category

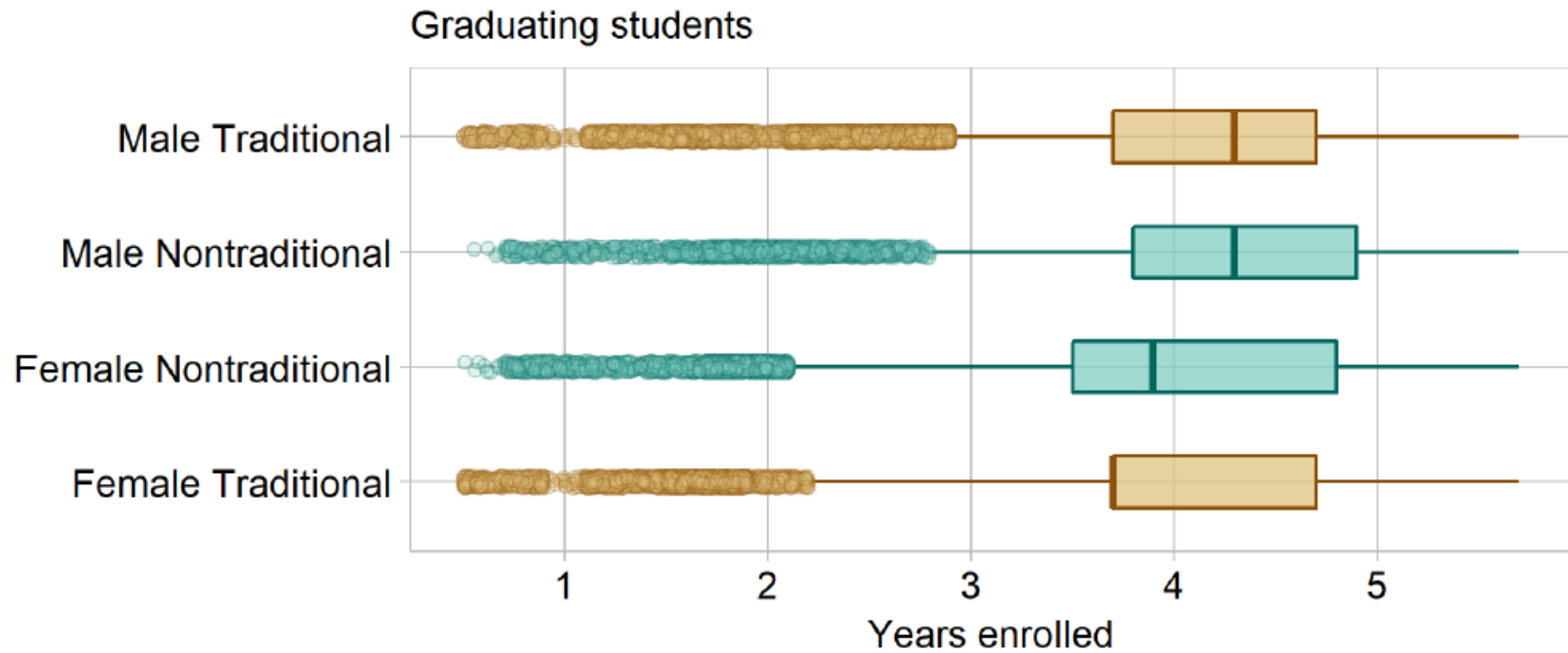


Data source (Unwin, 2015)

← quantitative variable →

# *box plot*

Years to graduation



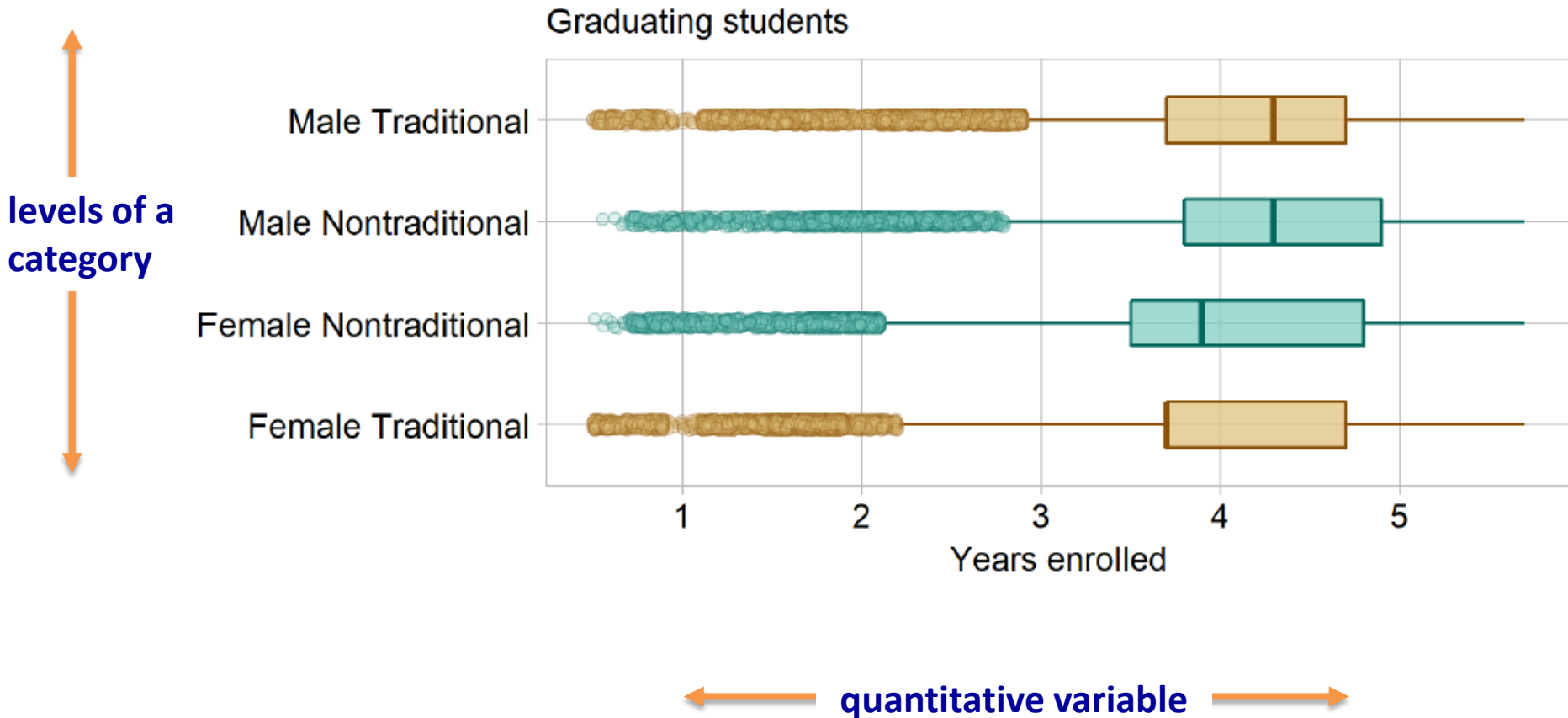


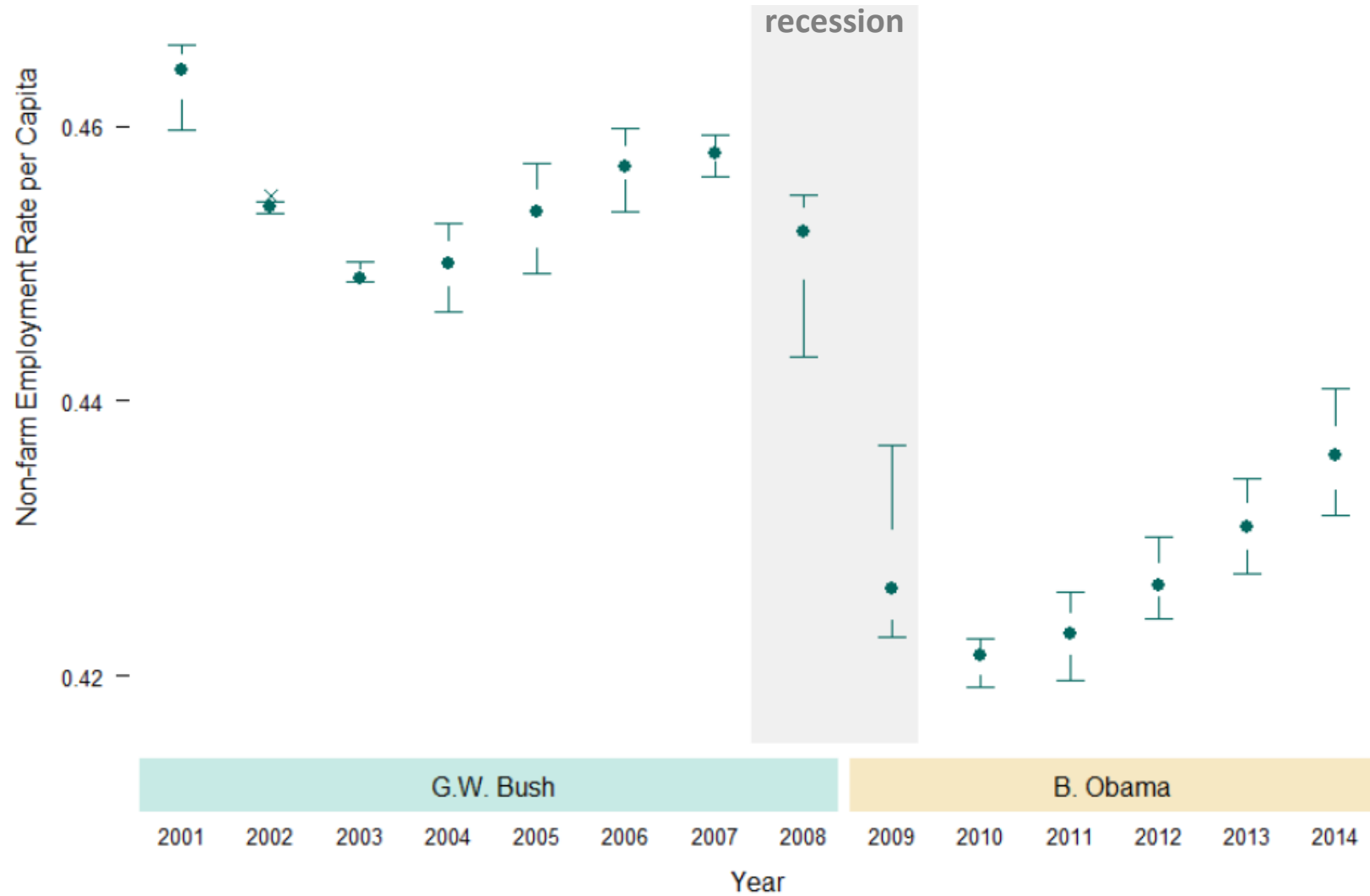
Story: **distribution and comparison**

Data: **1 quantitative, 1 categorical**

*box plot*

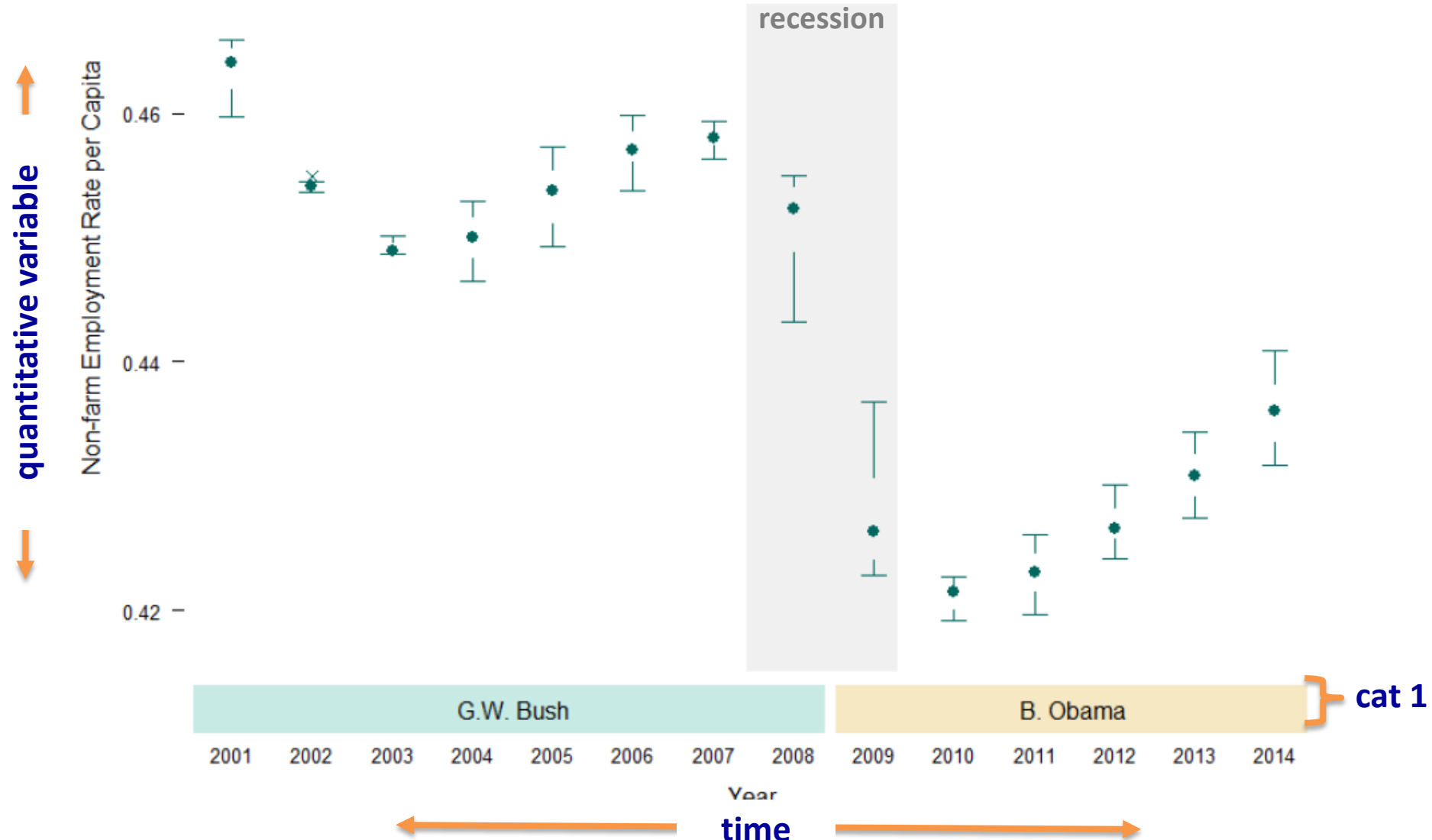
Years to graduation





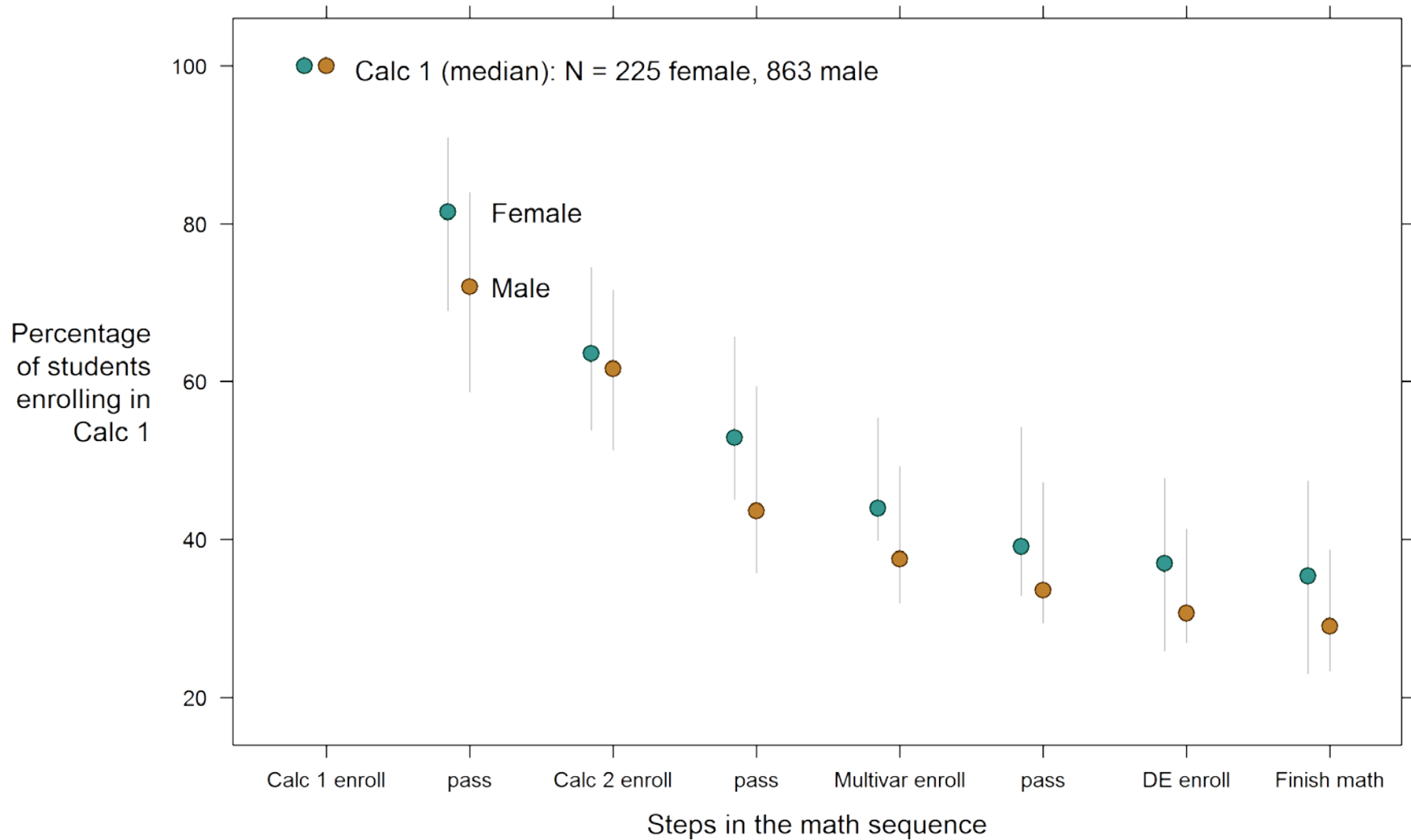
Story: evolution of distributions  
Data: 1 quantitative, 2 categorical

*box plot*  
US employment



# *distributions*

## Math sequence attrition



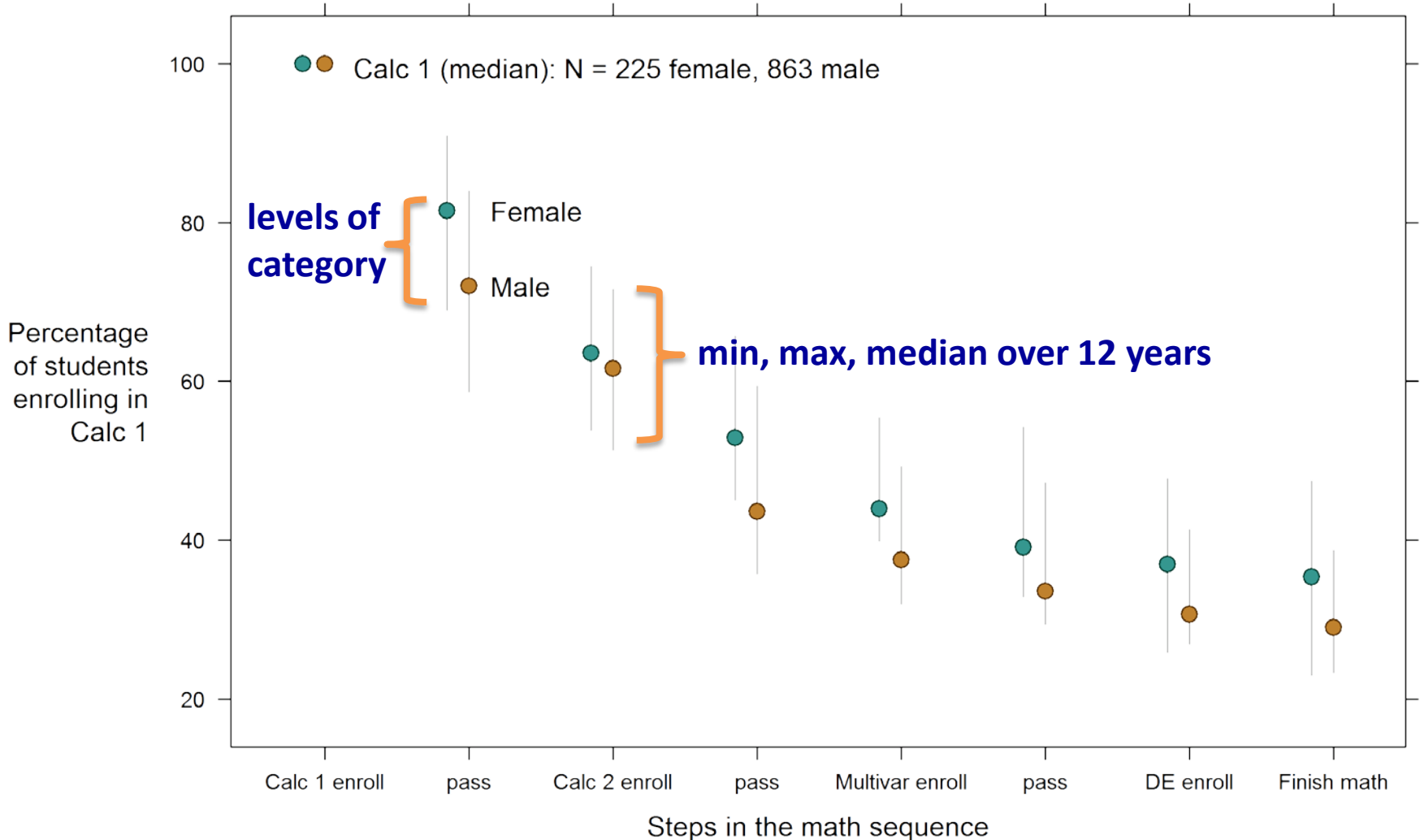
Story: evolution of distributions

Data: 1 quantitative, 2 categorical

*distributions*

Math sequence attrition

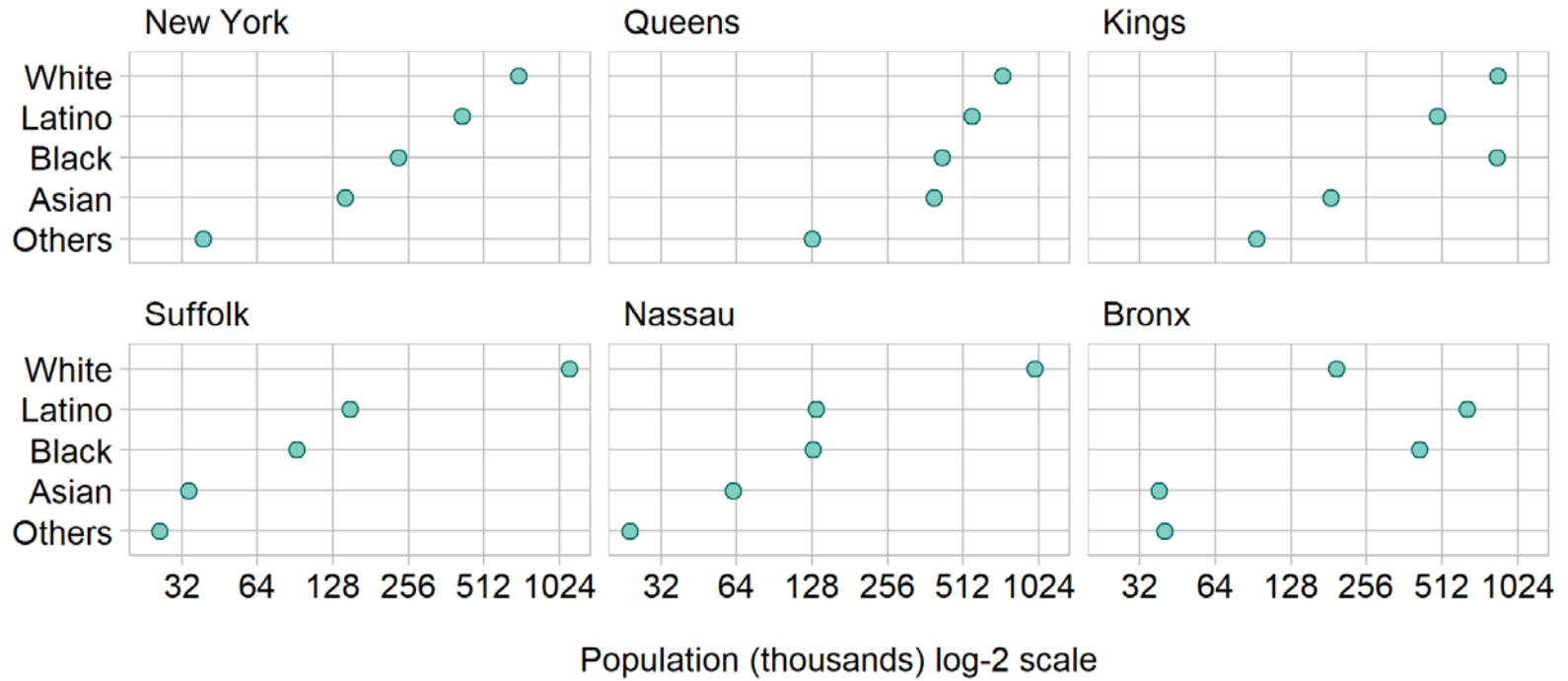
↑  
quantitative variable  
↓



← levels of category (evolution) →



## County population in NY State



Story: **comparing data**

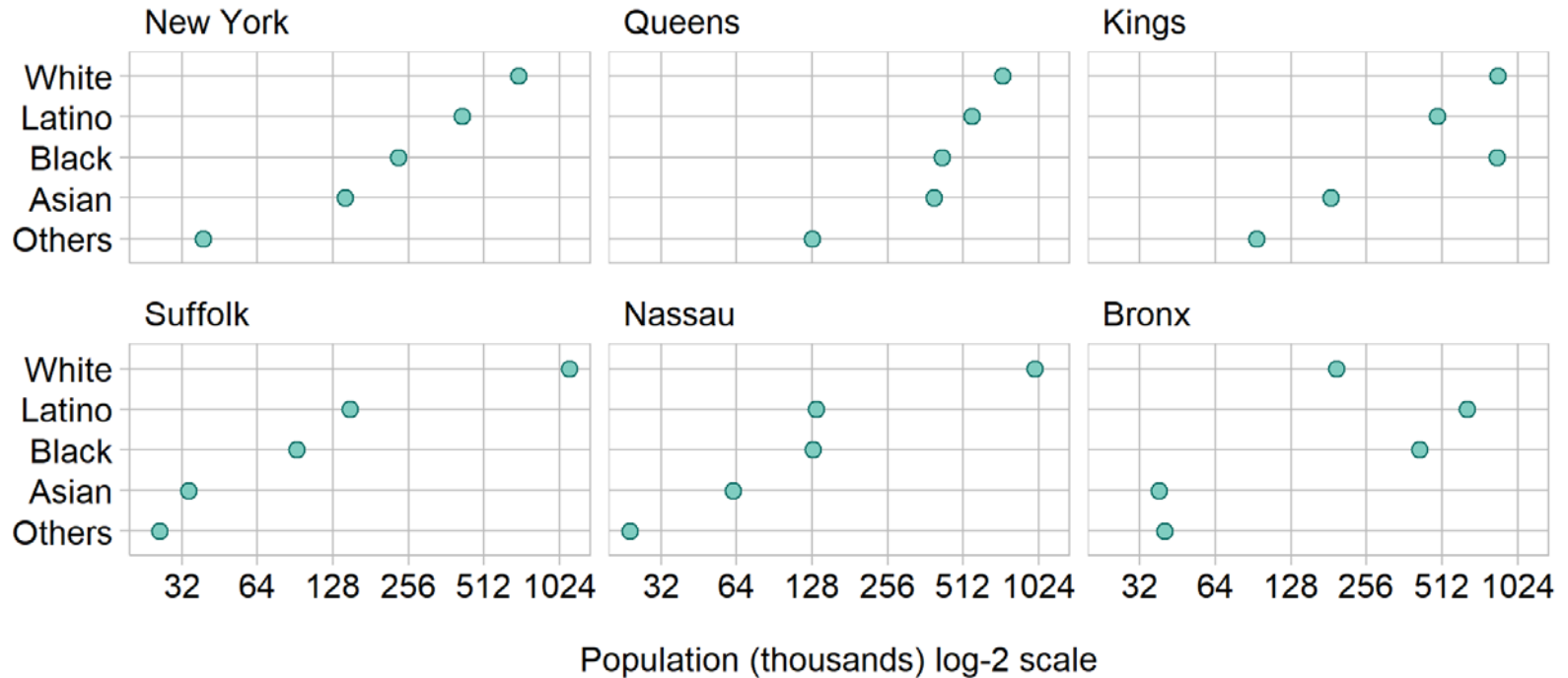
Data: **1 quantitative, 2 categorical**

*multiway*

County population in NY State

levels of a  
category

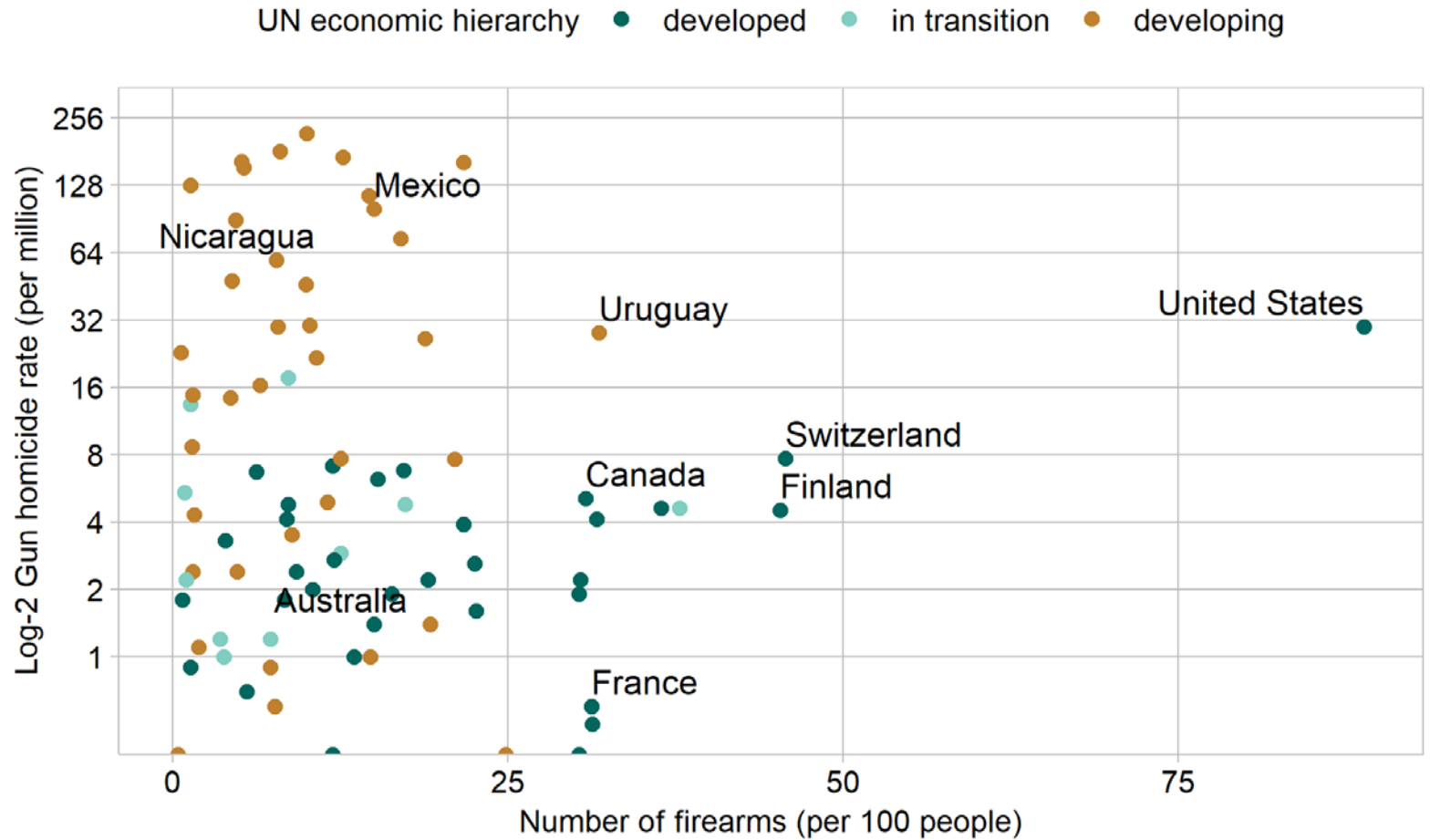
levels of a  
category



quantitative variable

# scatterplot

## Gun ownership and gun homicides



Story: comparing data and correlation

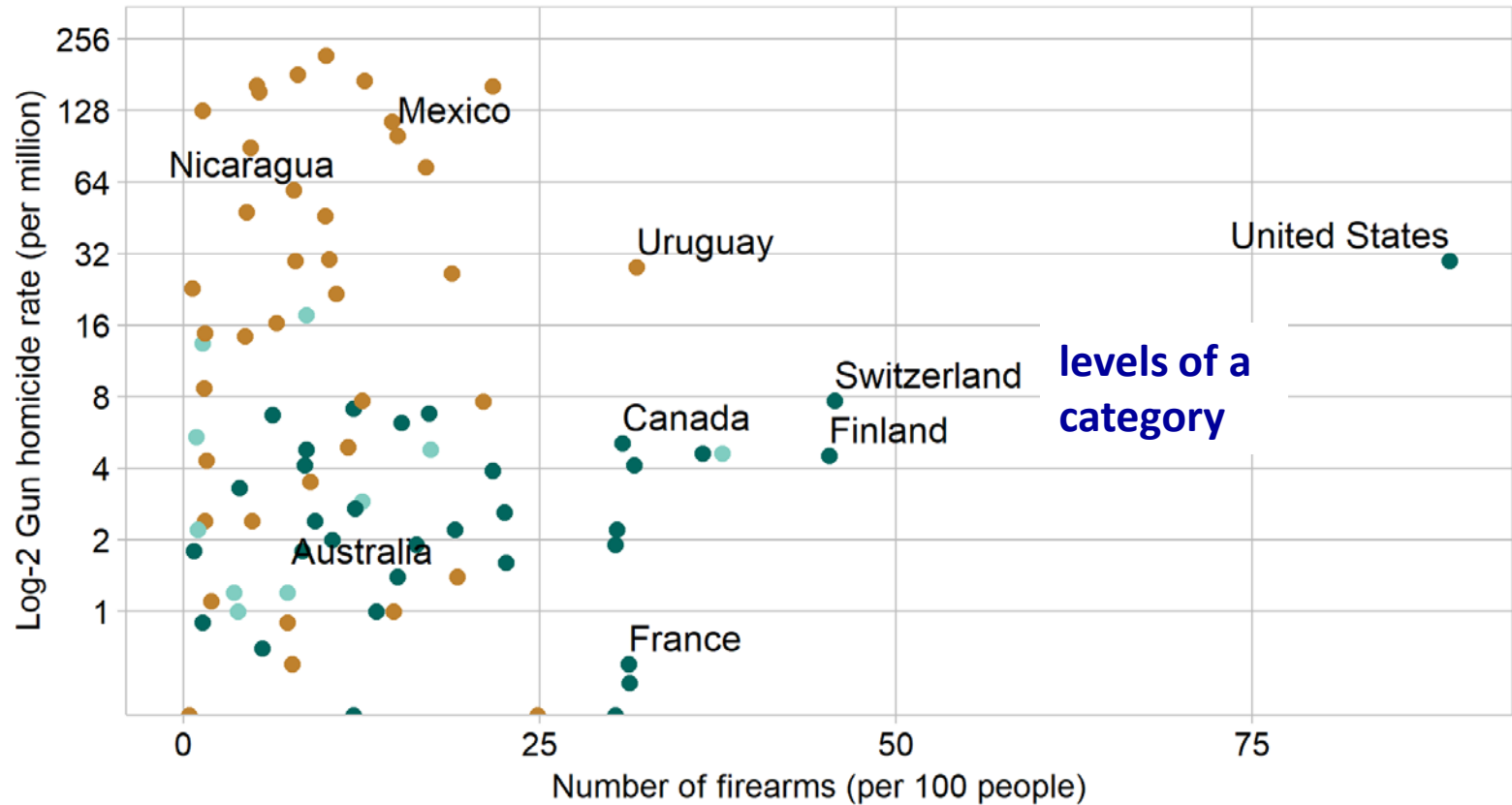
Data: 2 quantitative, 2 categorical

*scatterplot*

Gun ownership and gun homicides

levels of a category

UN economic hierarchy    ● developed    ● in transition    ● developing



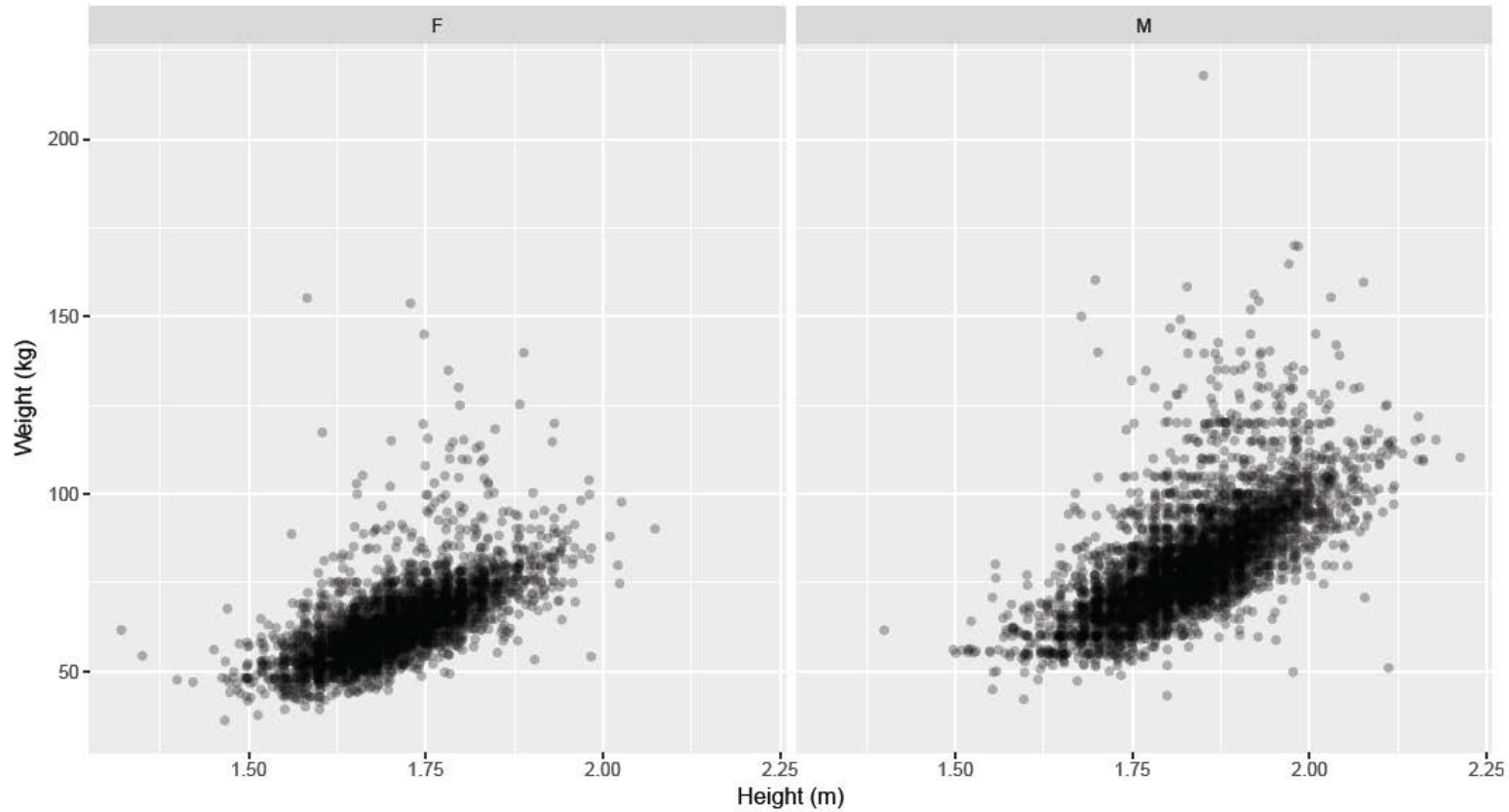
levels of a category

quantitative variable

# *scatterplot*

## *multi-panel*

Olympians' height and weight

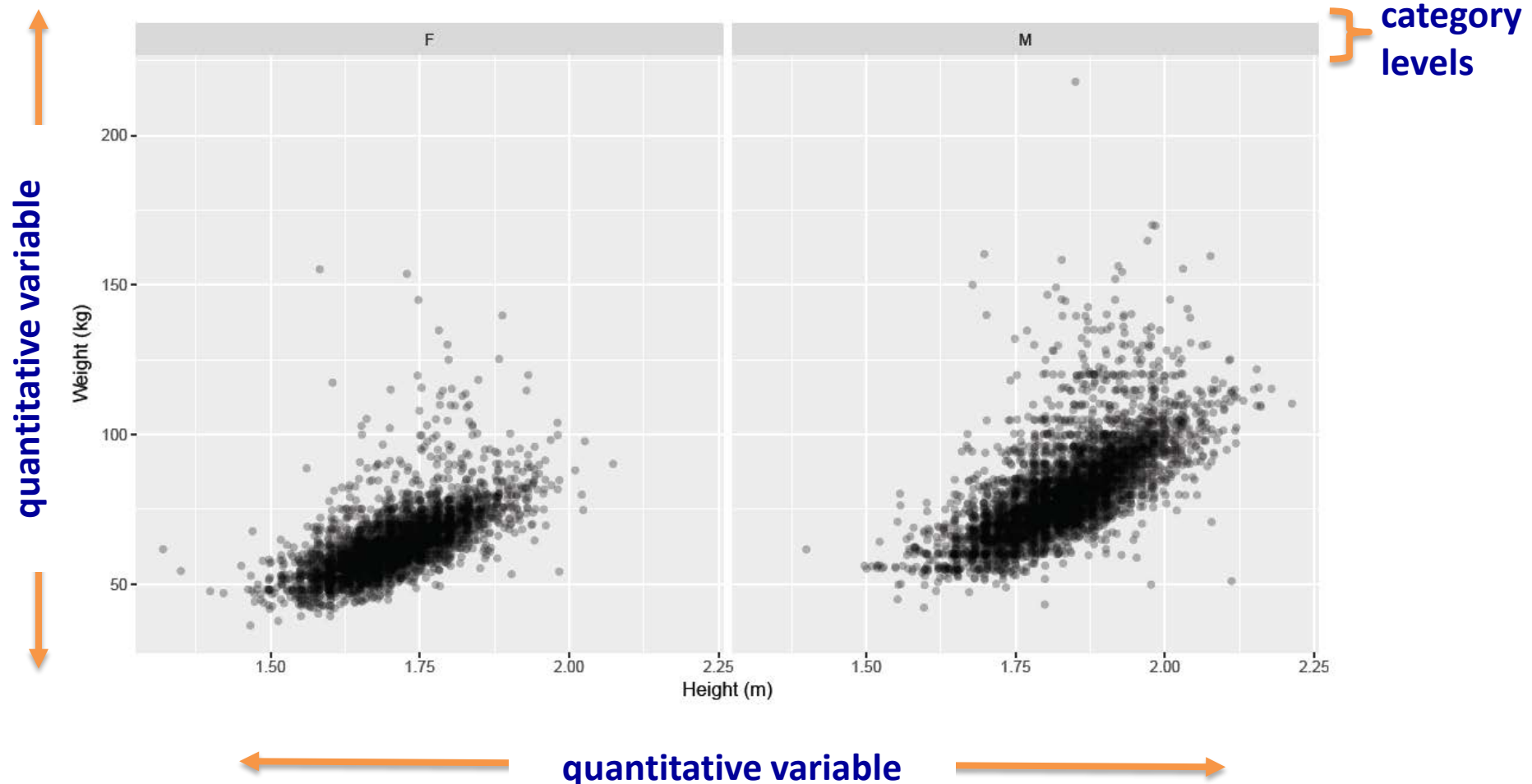




**Story:** correlation & comparison  
**Data:** 2 quantitative, 1 categorical

*scatterplot*  
*multi-panel*

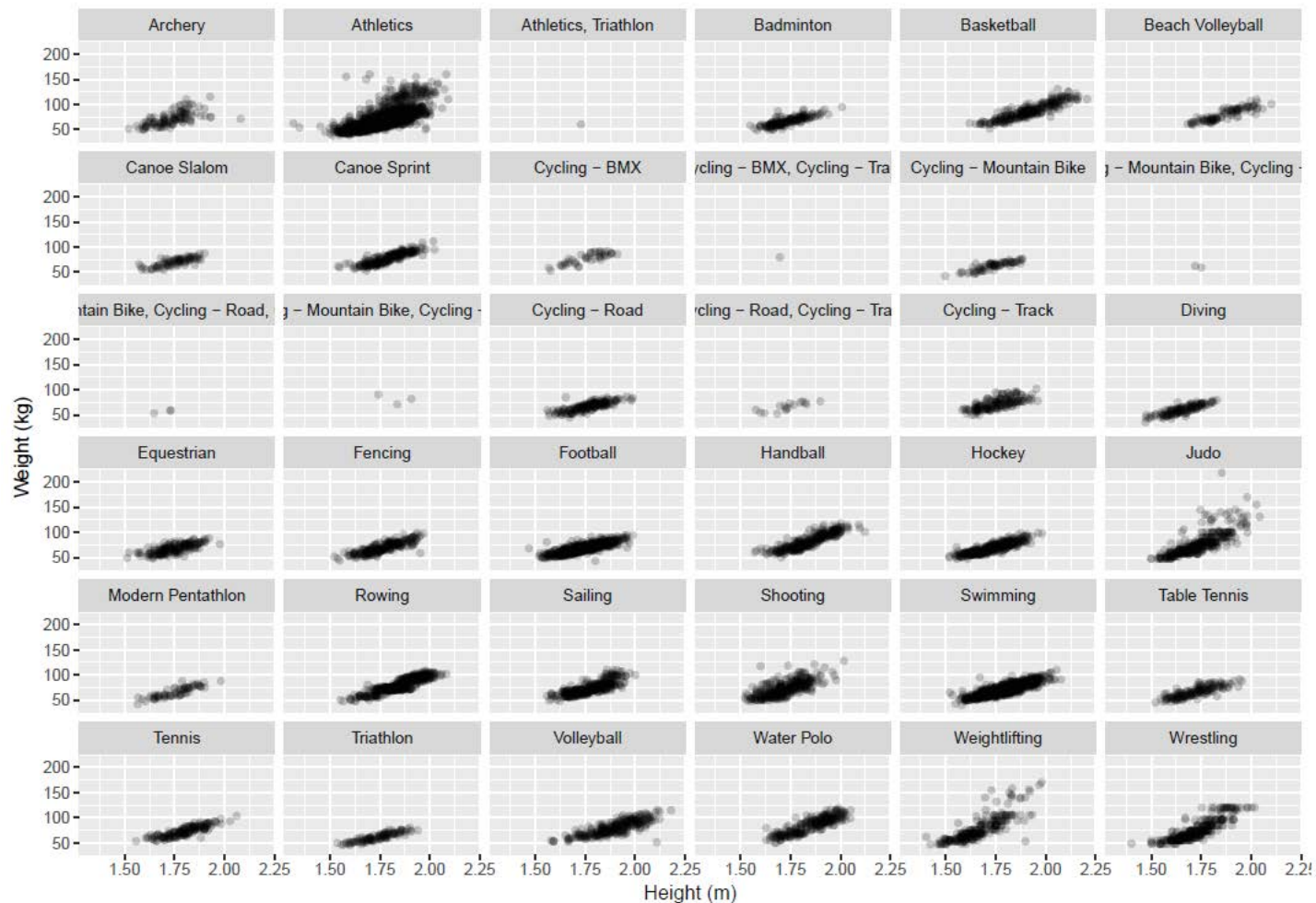
Olympians' height and weight



# *scatterplot*

## *small multiples*

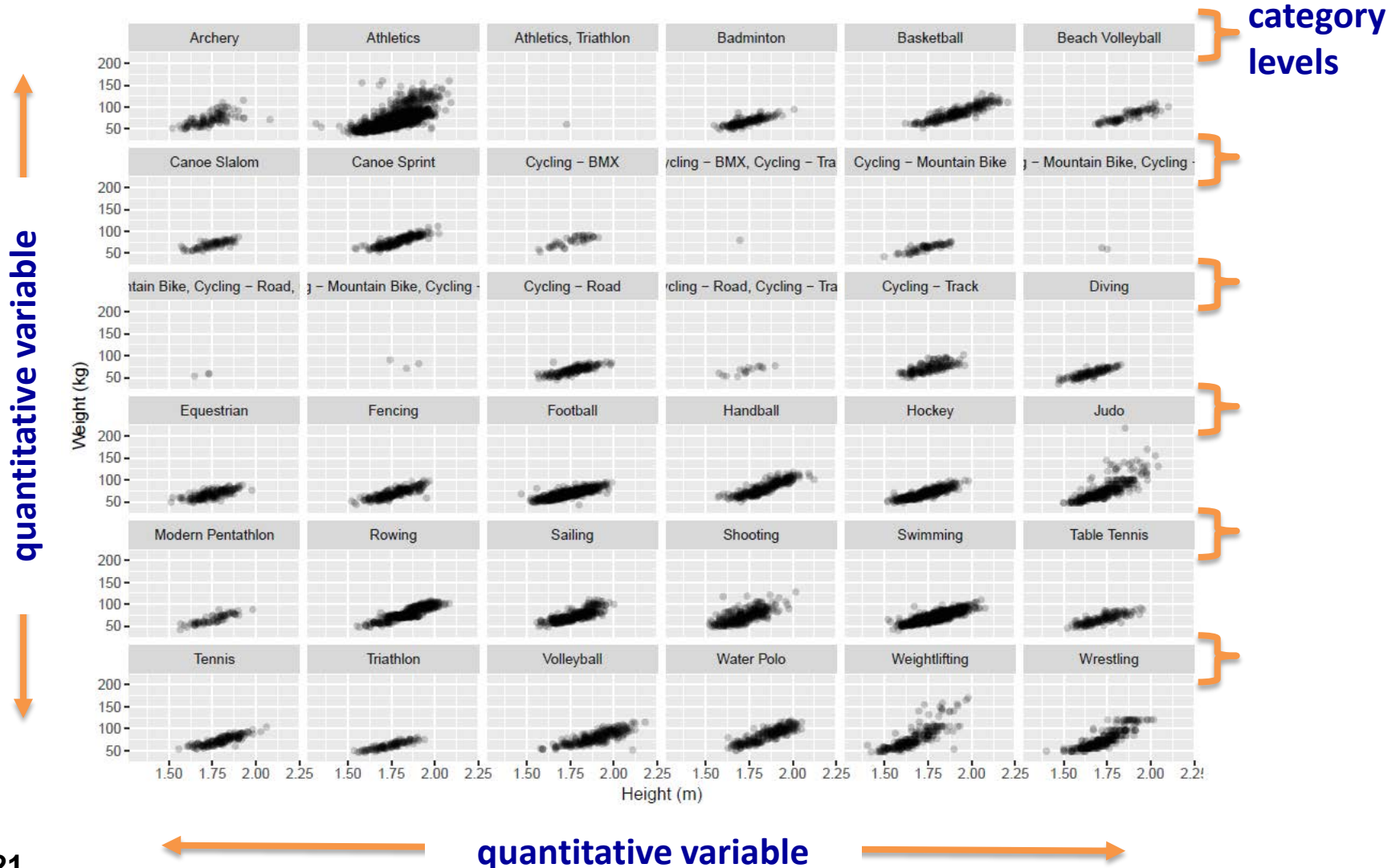
Olympians' height and weight, by event



Story: correlation & comparison  
Data: 2 quantitative, 1 categorical

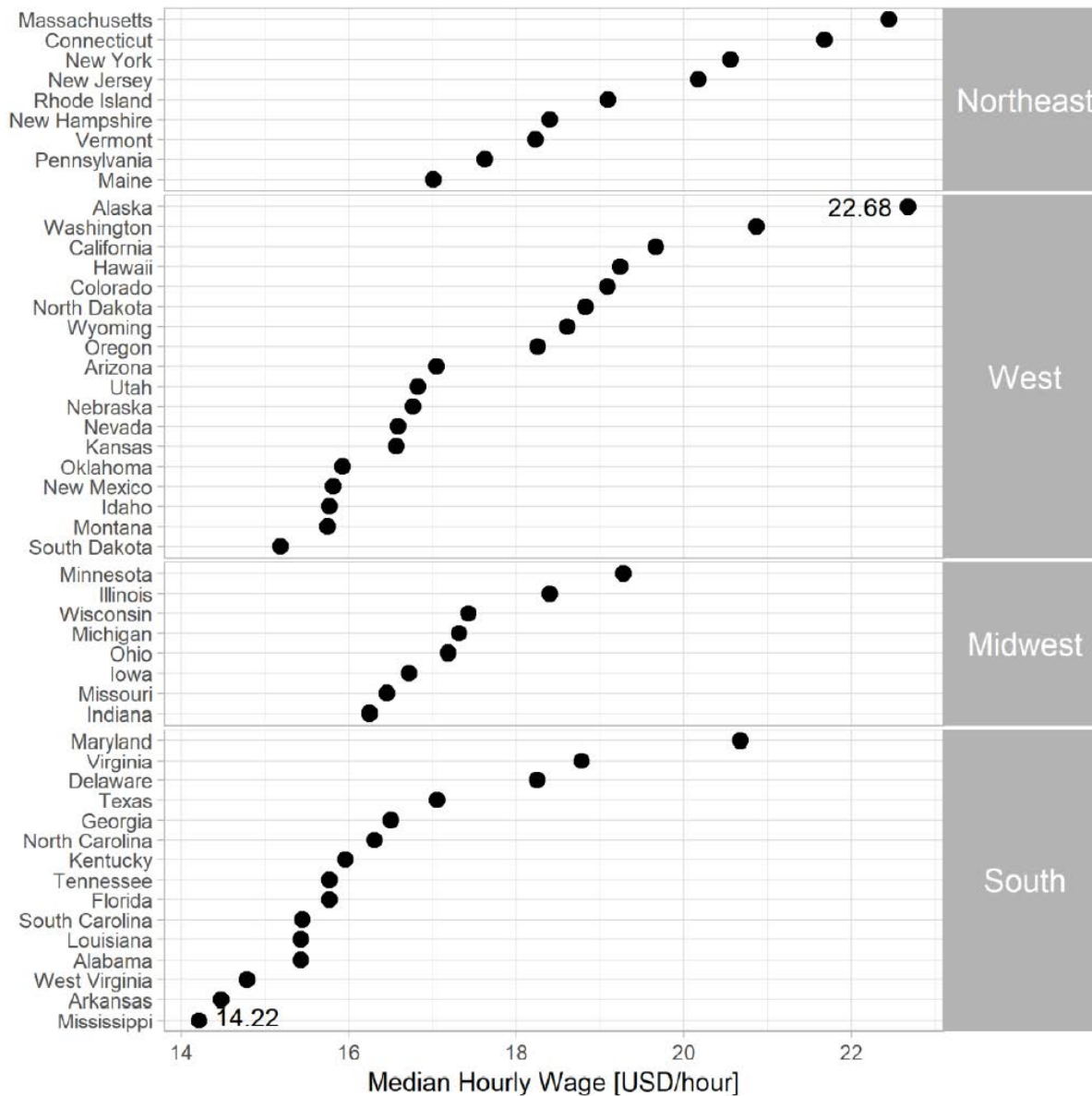
*scatterplot*  
*small multiples*

Olympians' height and weight, by event



# Cleveland dot plot

US median hourly wages



Story: **comparing data**

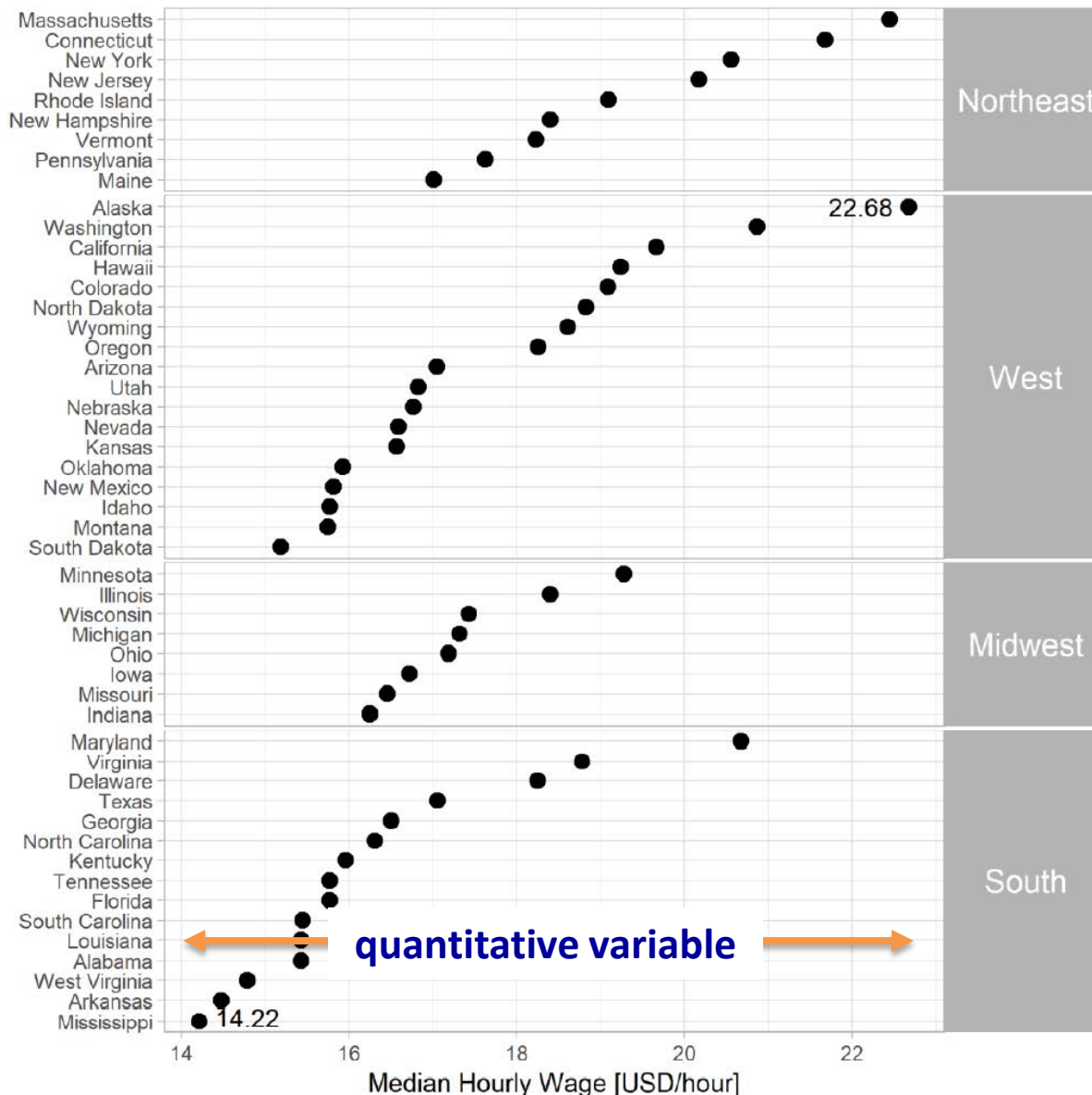
Data: **1 quantitative, 2 categorical**

# *Cleveland dot plot*

US median hourly wages

levels of a  
category

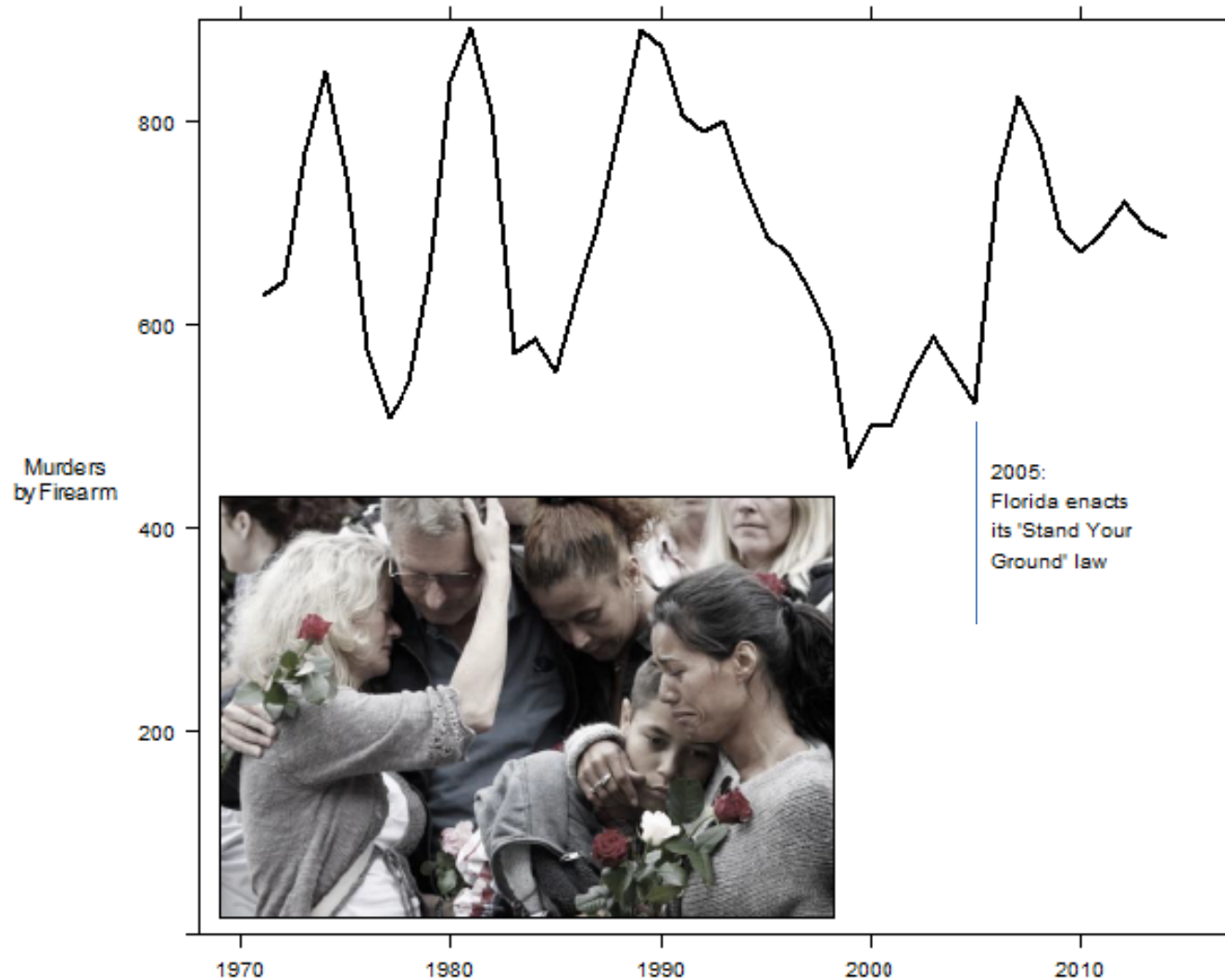
levels of a  
category





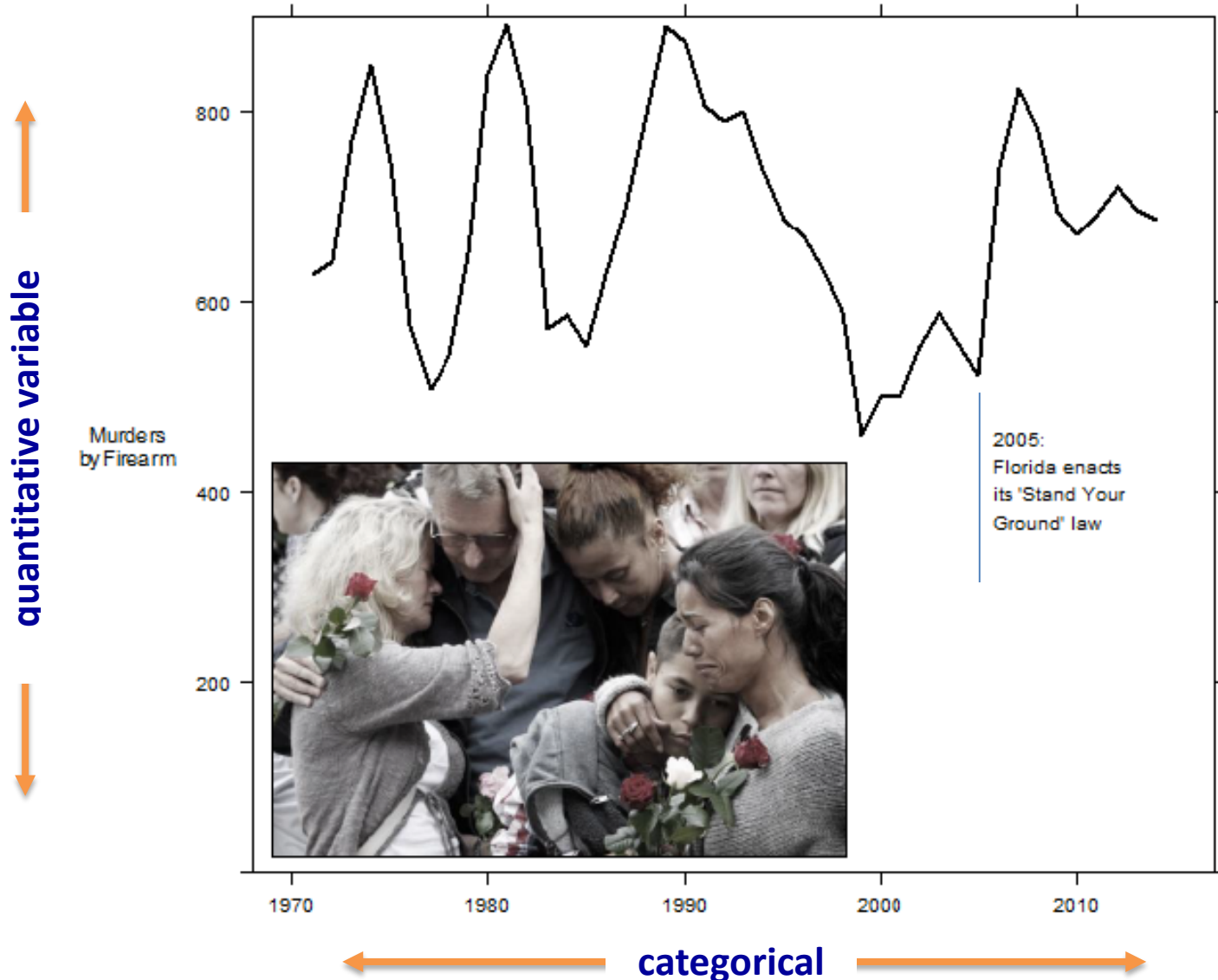
# *line graph*

US firearm deaths



Story: **evolution**  
Data: **1 quantitative, 1 categorical**

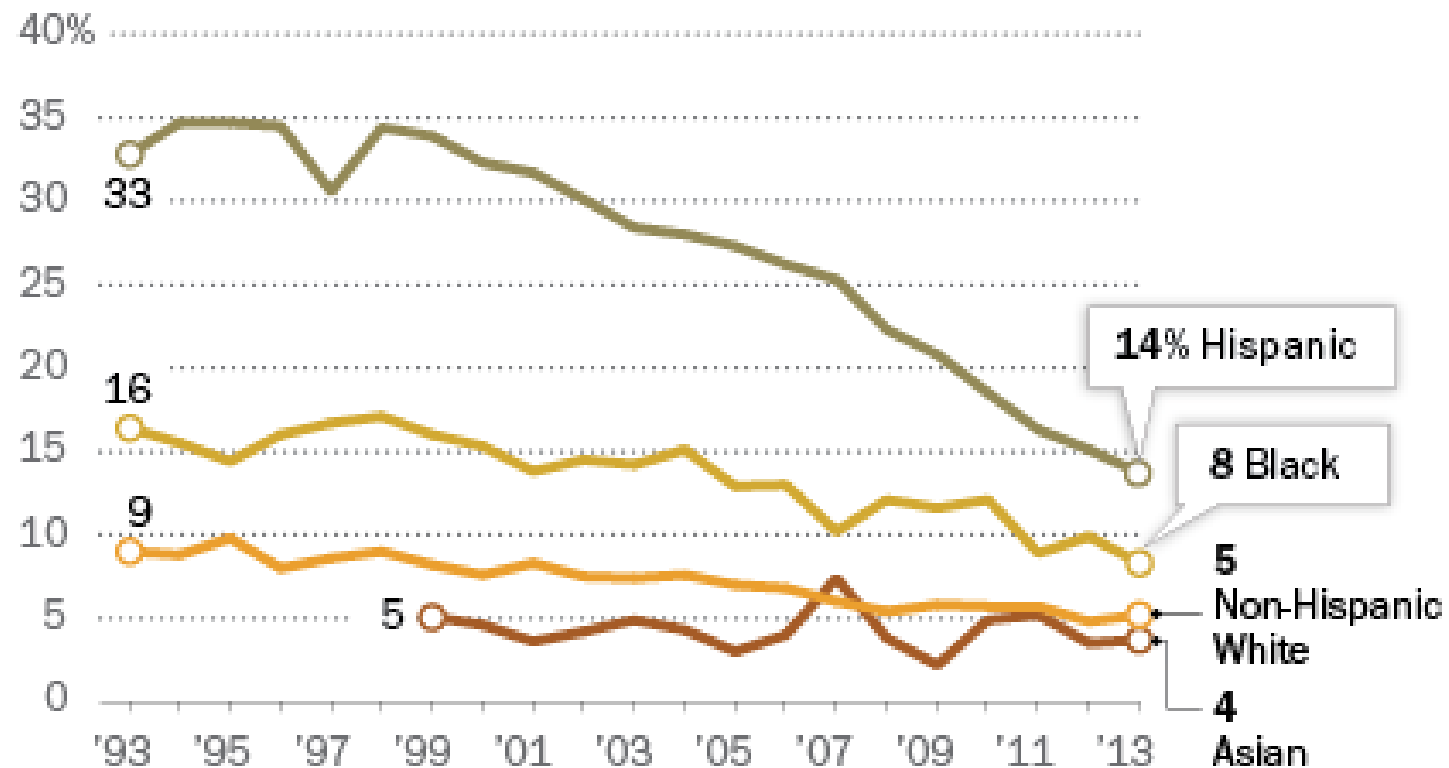
*line graph*  
US firearm deaths



# line graph

## US high school dropout rates

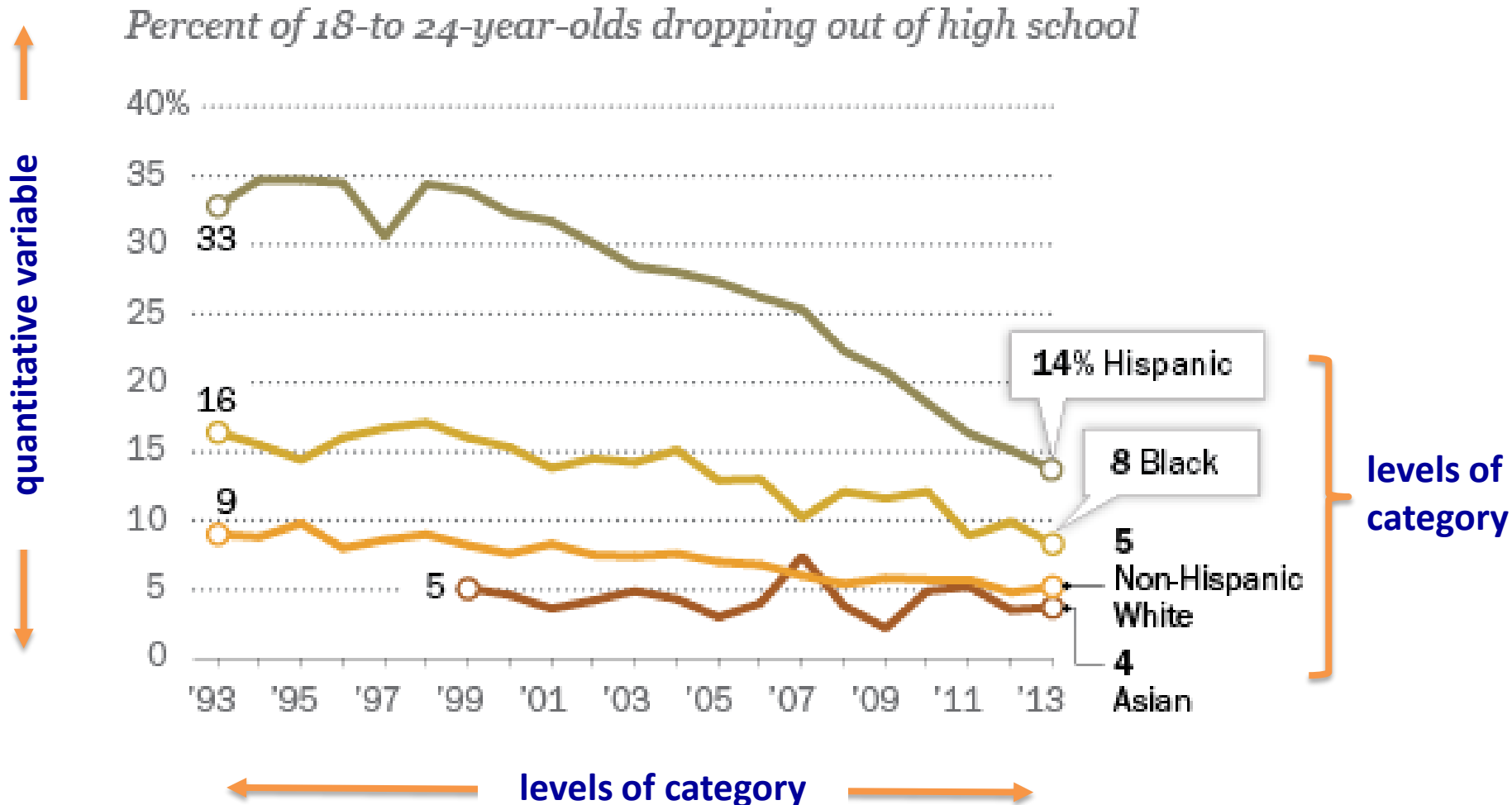
*Percent of 18-to 24-year-olds dropping out of high school*



Story: **evolution & comparison**  
Data: **1 quantitative, 2 categorical**

*line graph*

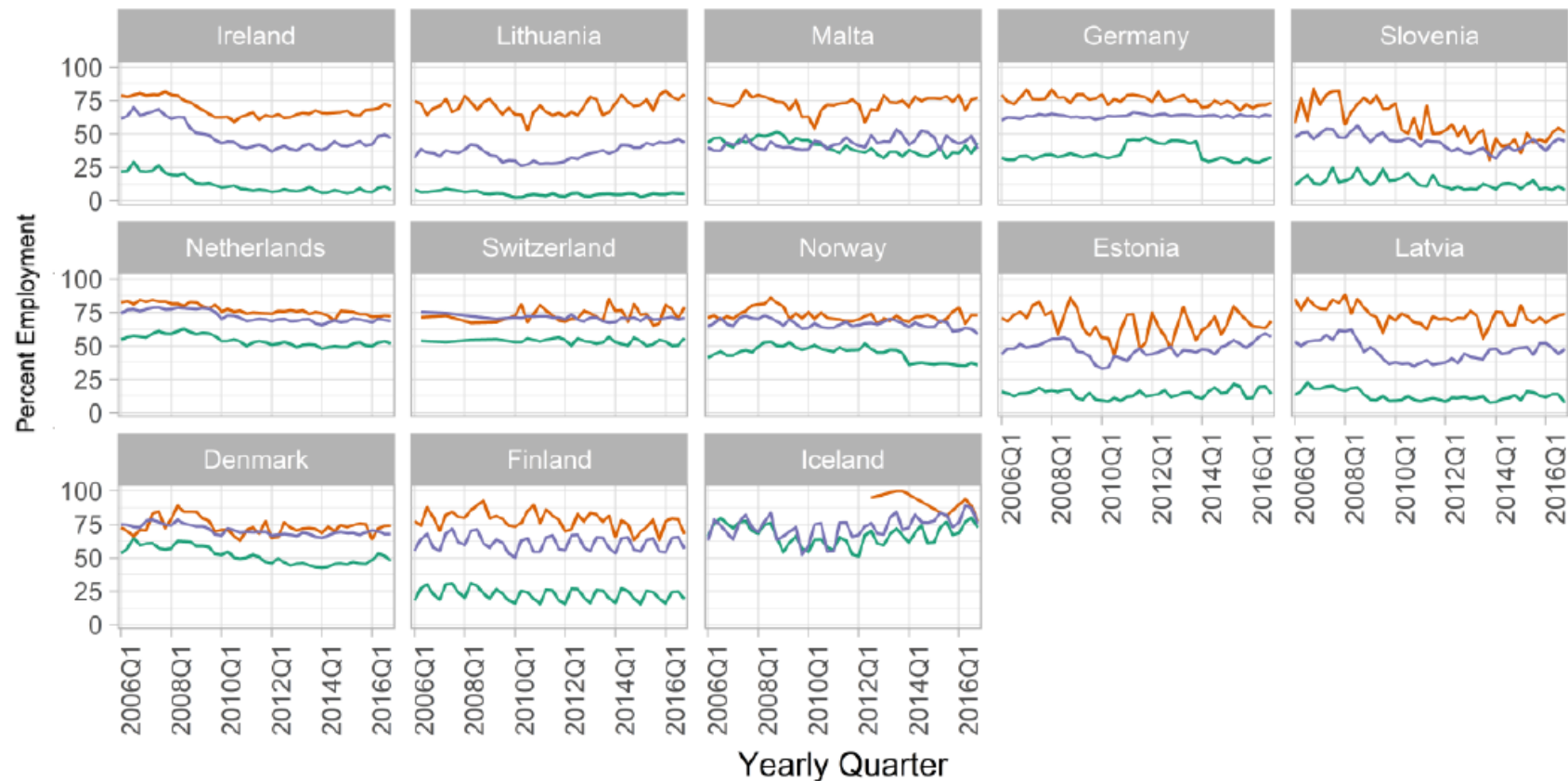
US high school dropout rates



# line graph

## small multiples

### EU employment by education level



#### Education Level Completed

- Less than primary, primary and lower secondary education (levels 0-2)
- Upper secondary and post-secondary non-tertiary education (levels 3 and 4)
- Tertiary education (levels 5-8)



Story: evolution and comparing data

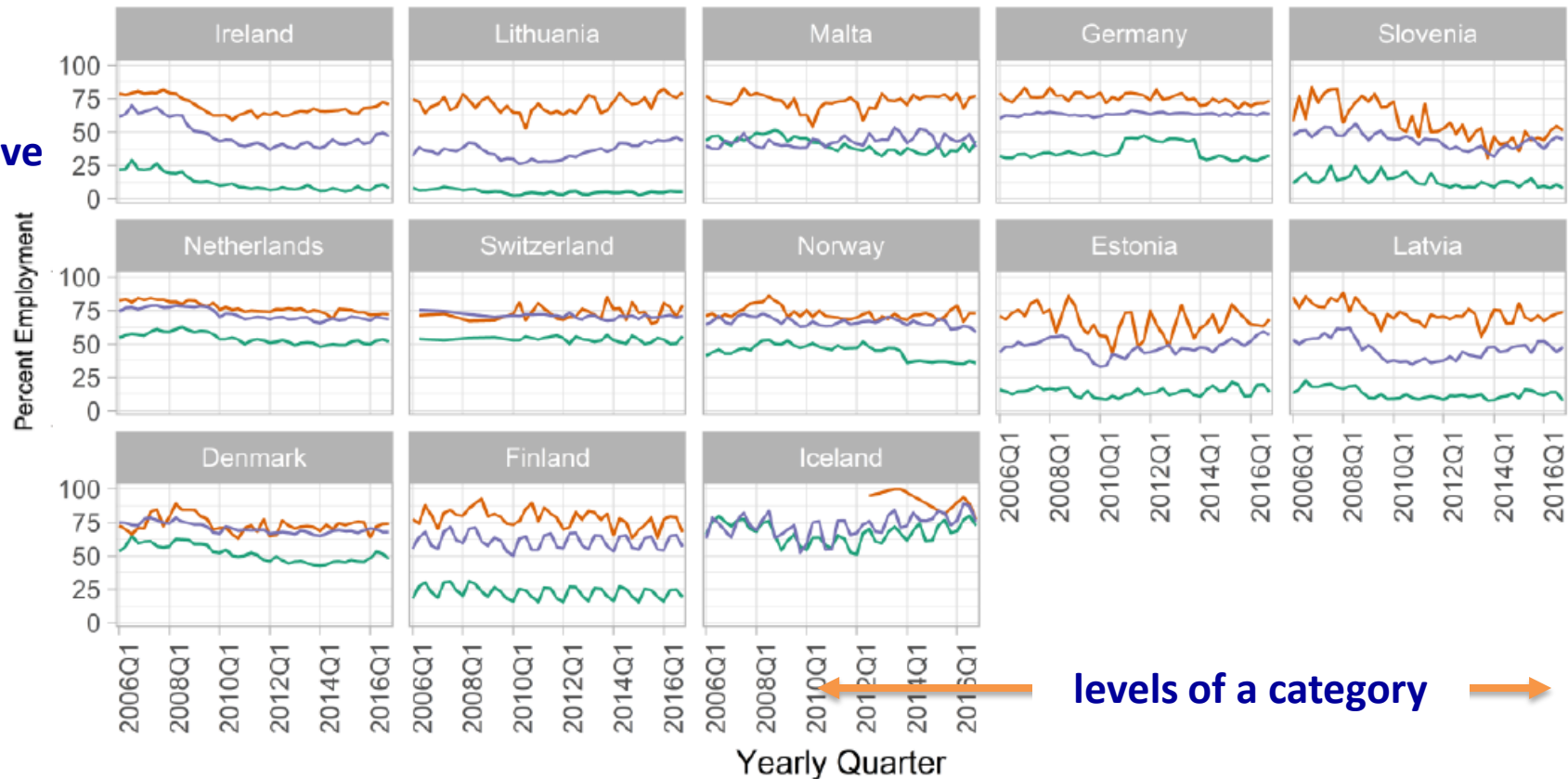
Data: 1 quantitative, 3 categorical

*line graph*

*small multiples*

EU employment by education level

levels of a category



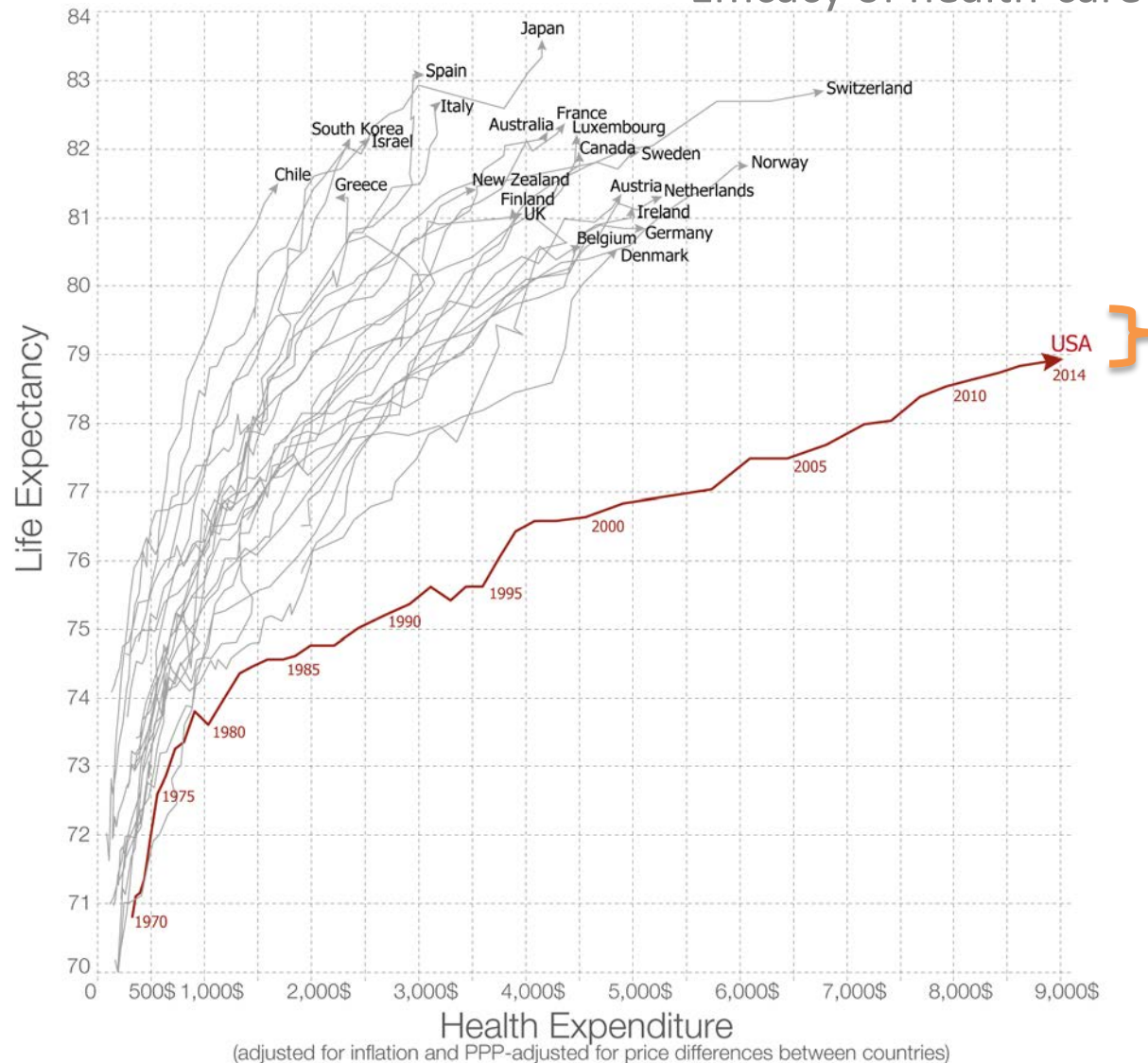
Education Level Completed

- Less than primary, primary and lower secondary education (levels 0-2)
- Upper secondary and post-secondary non-tertiary education (levels 3 and 4)
- Tertiary education (levels 5-8)

levels of a category

# line graph

## Efficacy of health-care expenditures

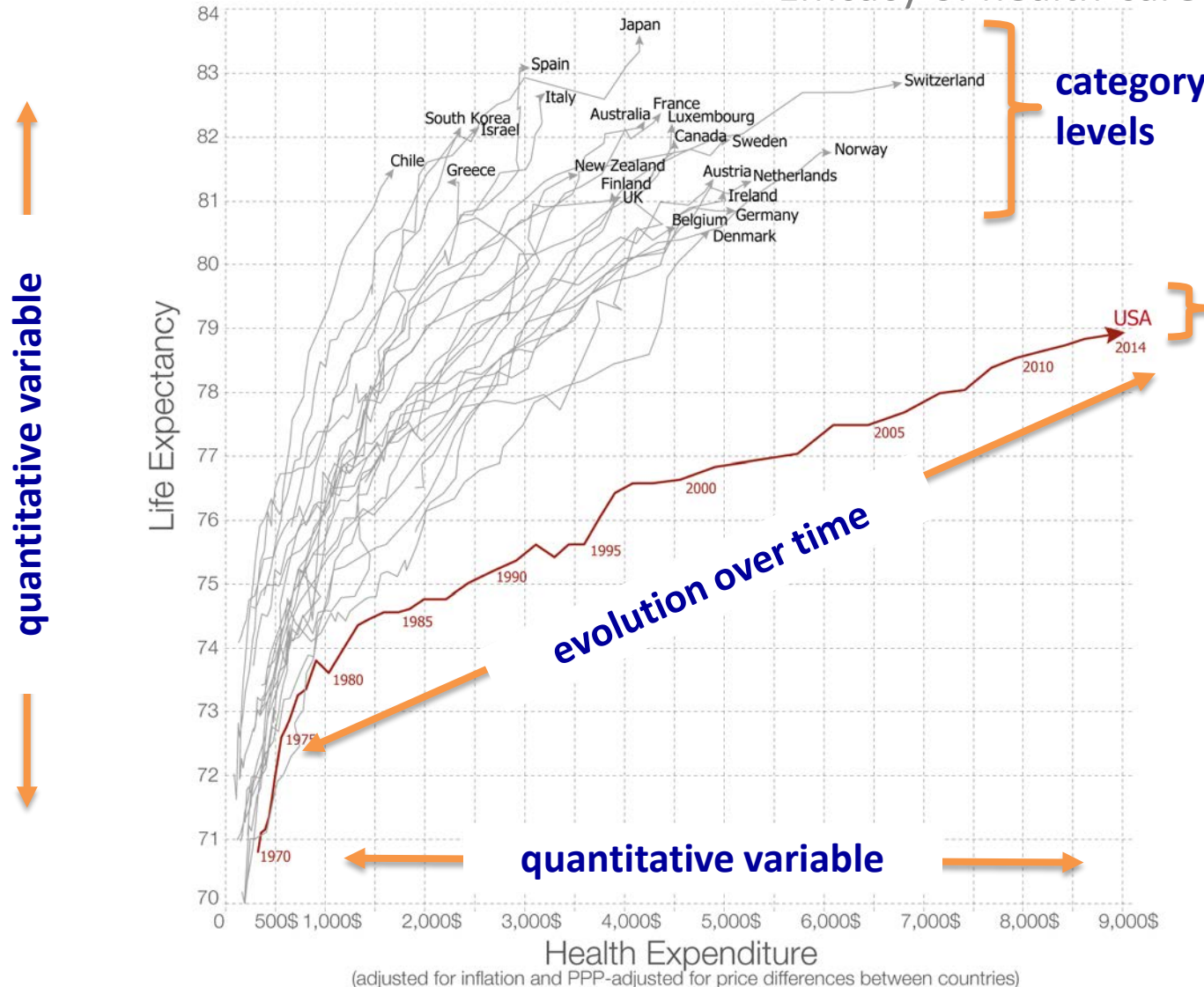


Data source: Health expenditure from the OECD; Life expectancy from the World Bank. Licensed under CC-BY-SA by the author Max Roser.  
The interactive data visualization is available at [OurWorldinData.org](http://OurWorldinData.org). There you find the raw data and more visualizations on this topic.

Story: correlation & comparison  
Data: 2 quantitative, 2 categorical

*line graph*

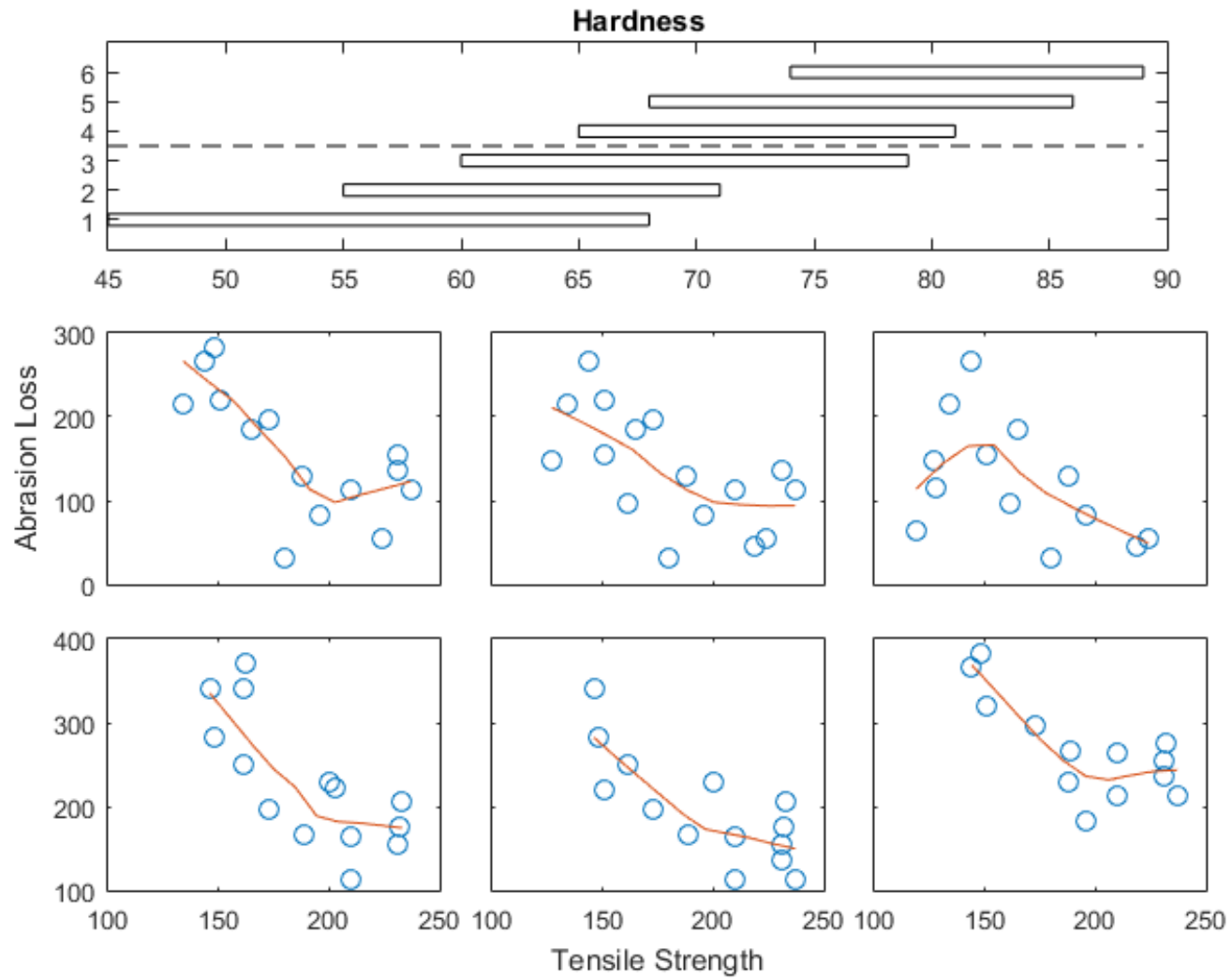
Efficacy of health-care expenditures



Data source: Health expenditure from the OECD; Life expectancy from the World Bank. Licensed under CC-BY-SA by the author Max Roser.  
The interactive data visualization is available at [OurWorldinData.org](https://www.ourworldindata.org). There you find the raw data and more visualizations on this topic.

# *conditioning plot*

Properties of rubber

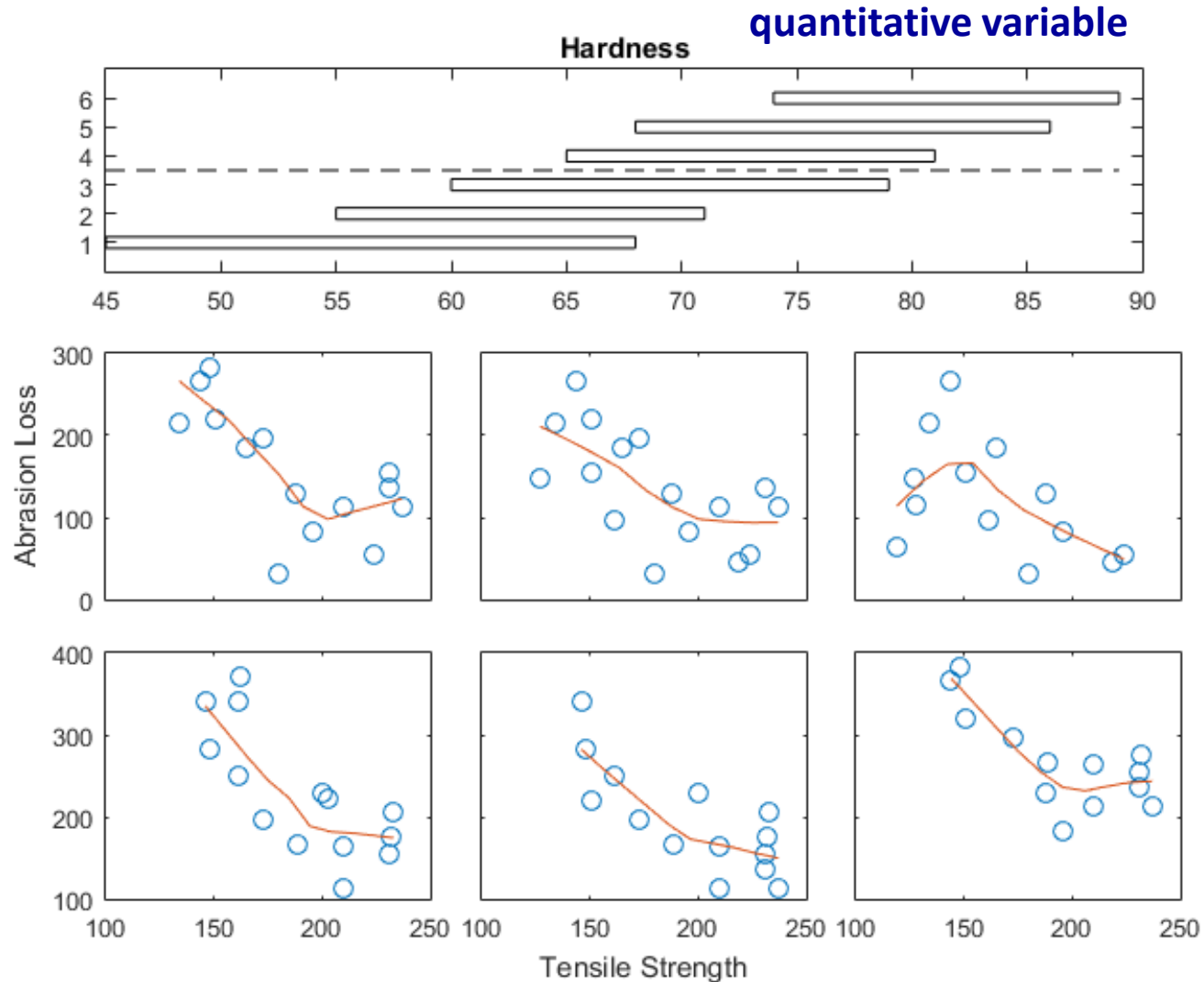


Story: **revealing correlations**

Data: **3 quantitative**

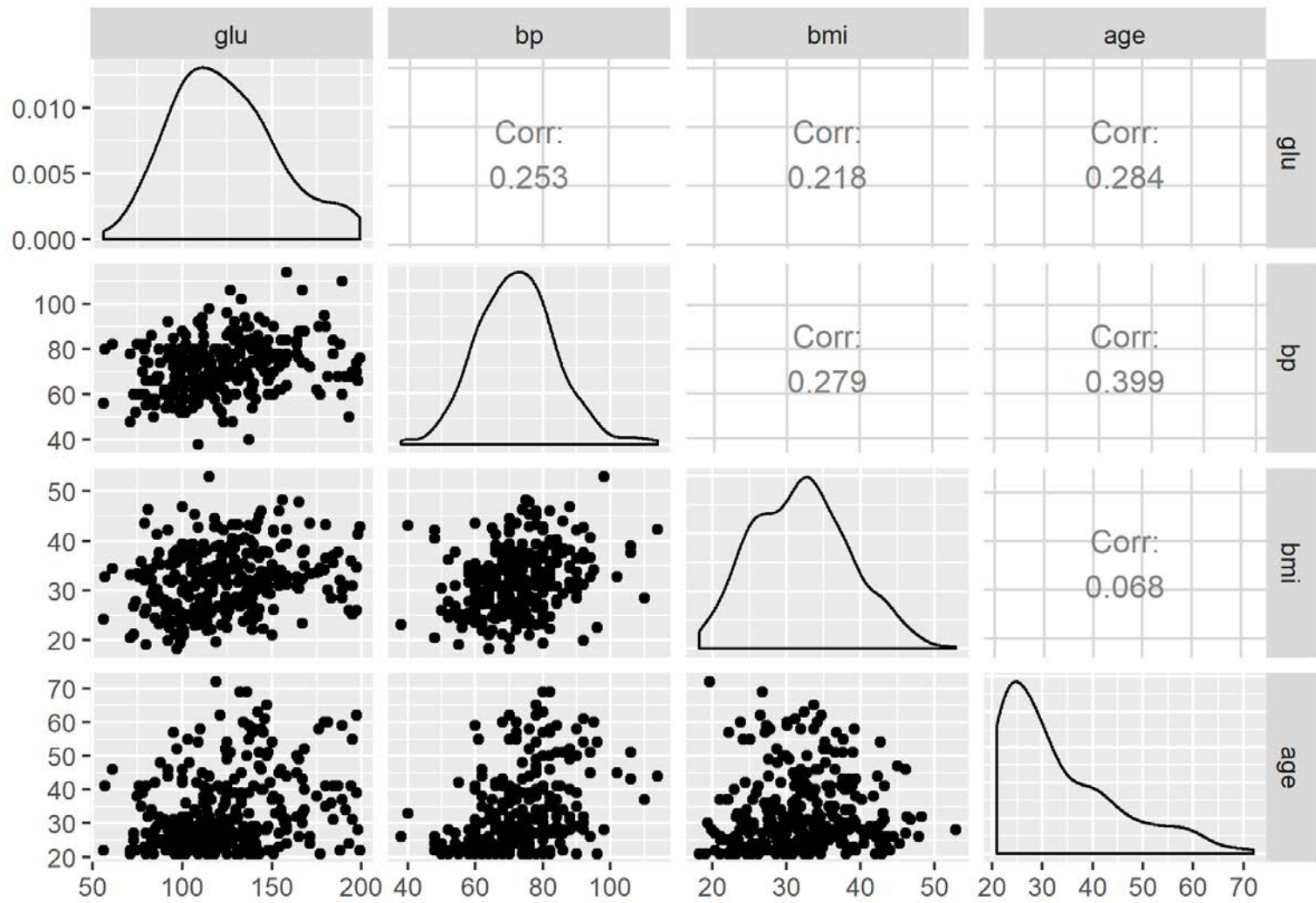
*conditioning plot*

Properties of rubber



# *scatterplot matrix*

Diabetes in Pima women



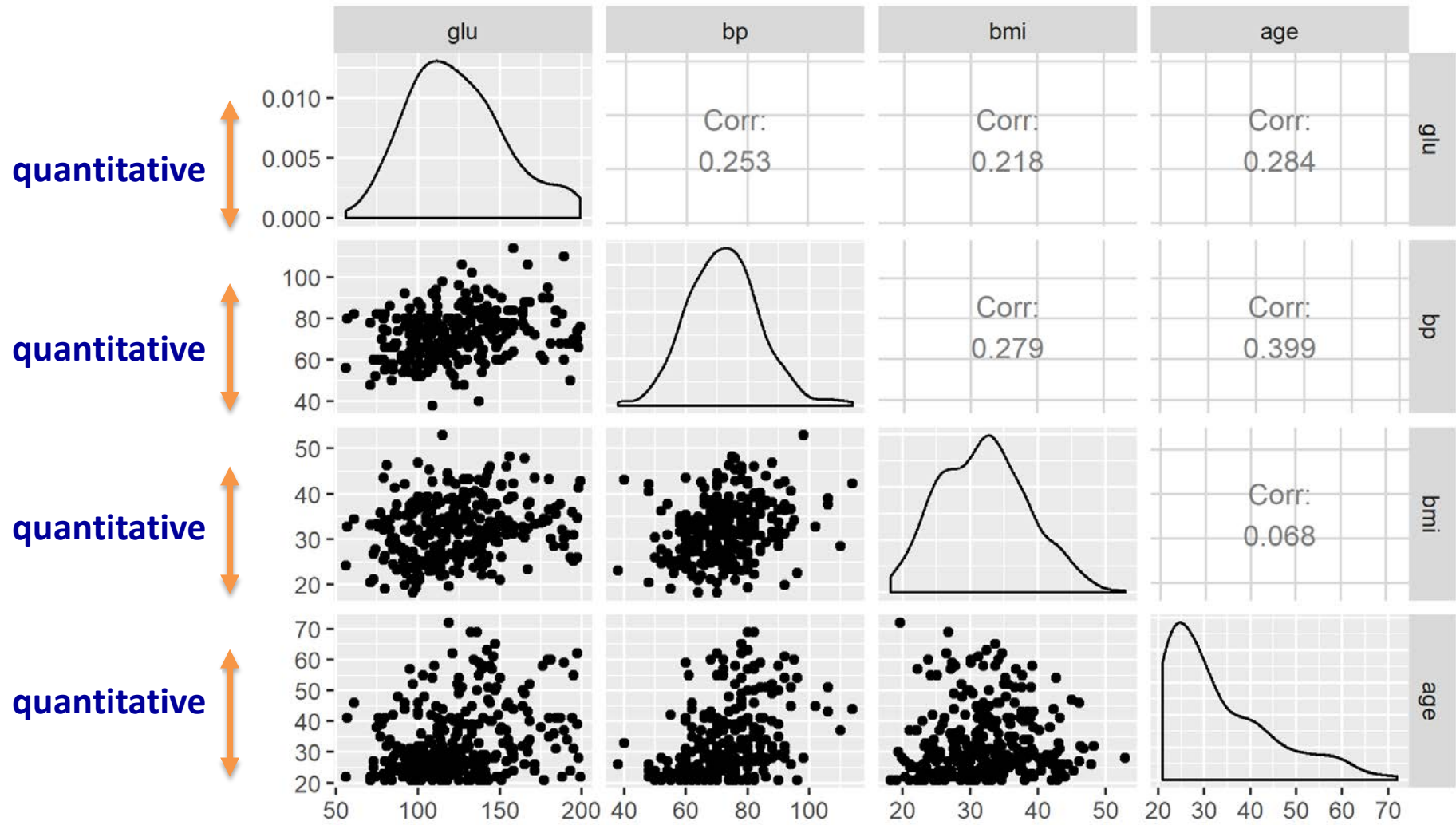


Story: revealing correlations

Data: 4 quantitative

# *scatterplot matrix*

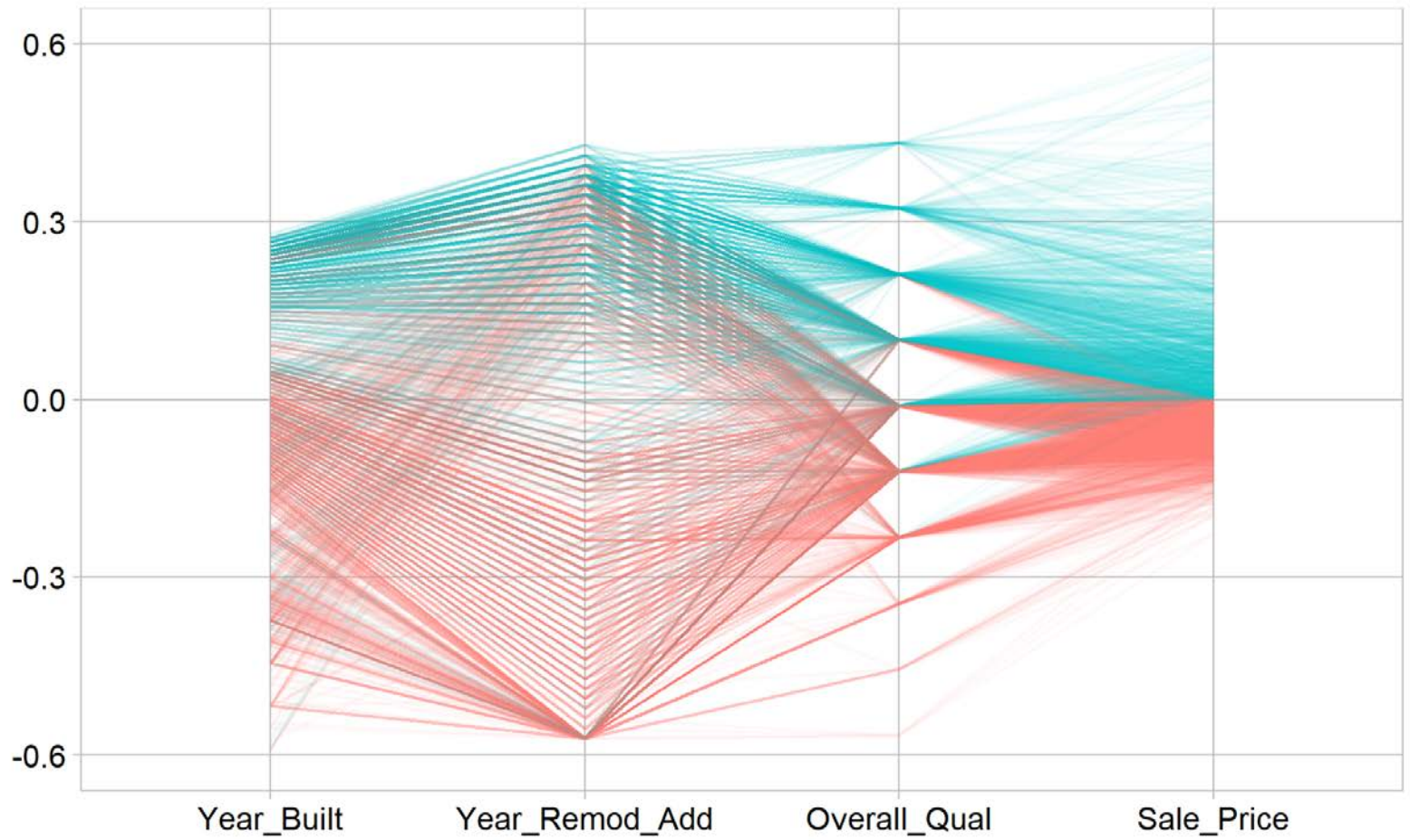
Diabetes in Pima women





# *parallel coordinate*

House pricing

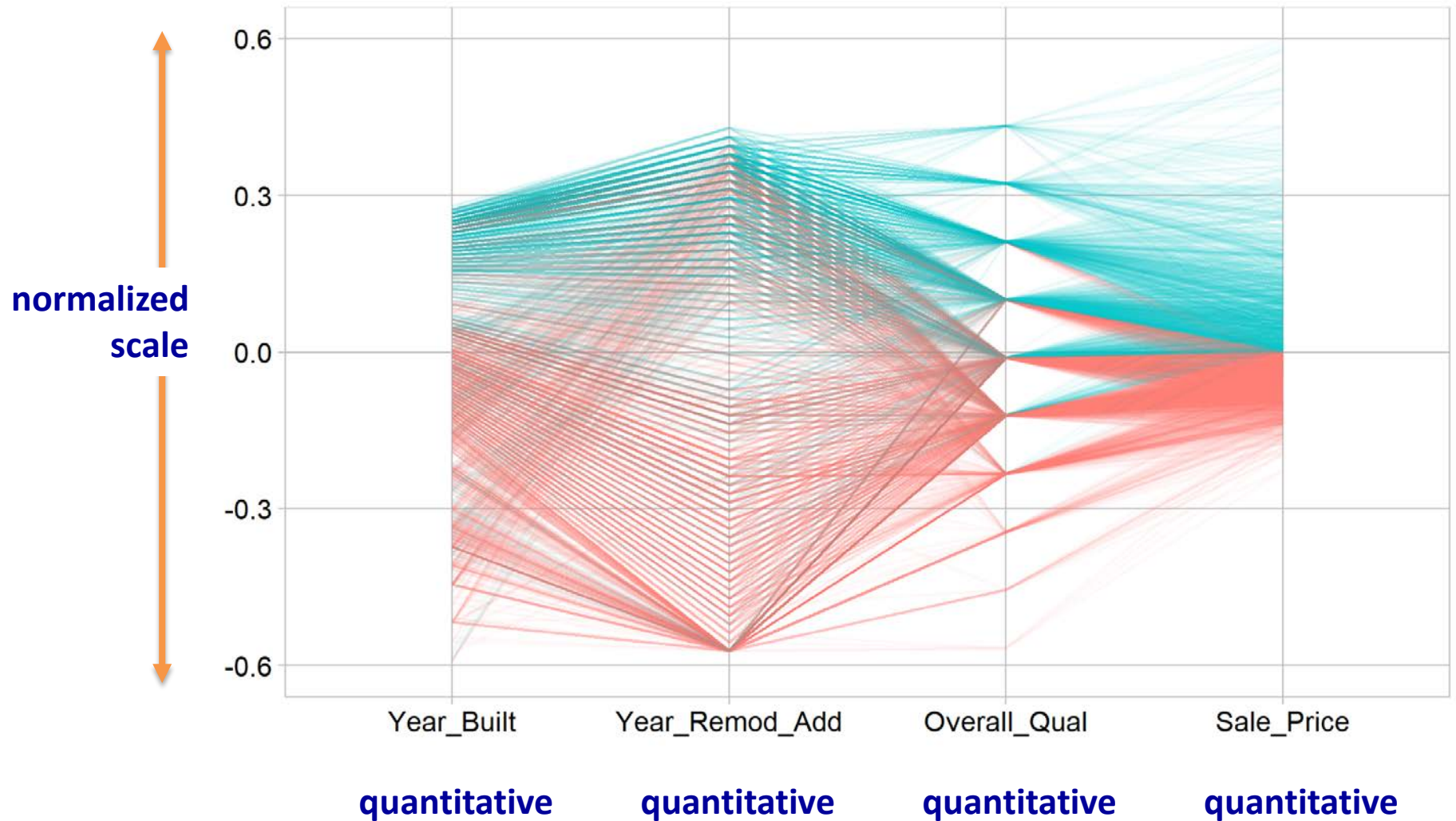


Story: **revealing correlations**

Data: **4 quantitative**

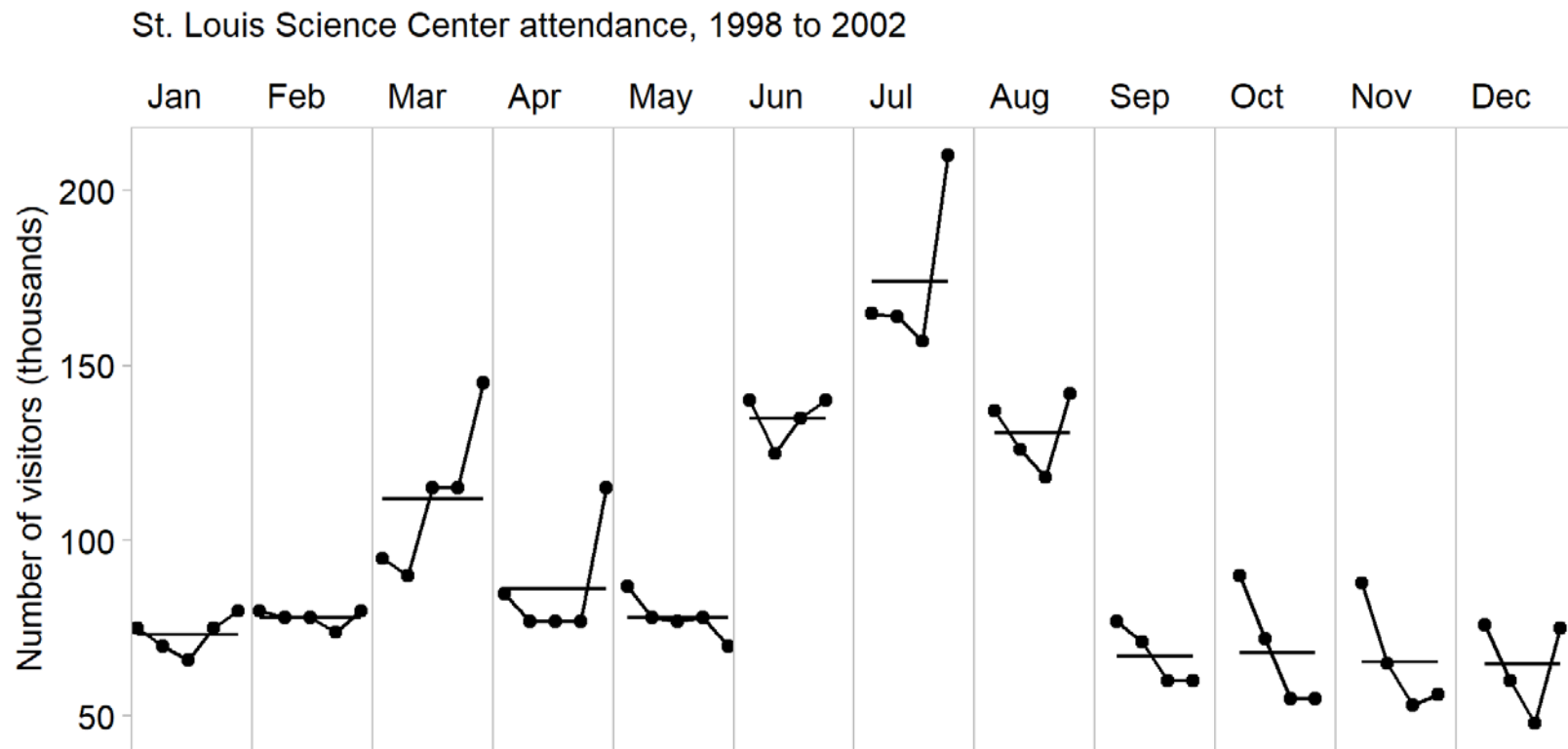
*parallel coordinate*

House pricing



# *cycle plot*

Science center attendance



**Story:** evolution and comparison  
**Data:** 1 quantitative, 2 categorical

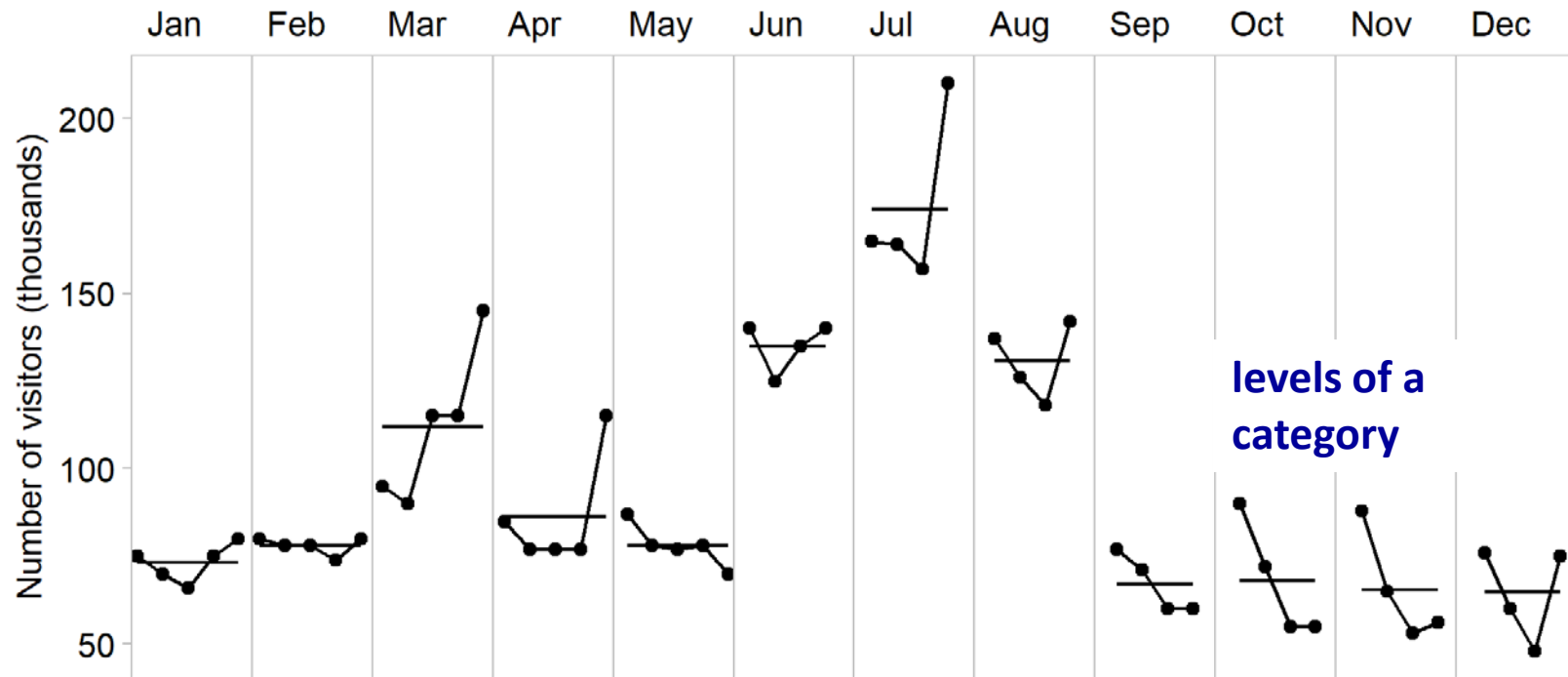
*cycle plot*

Science center attendance

← categorical variable →

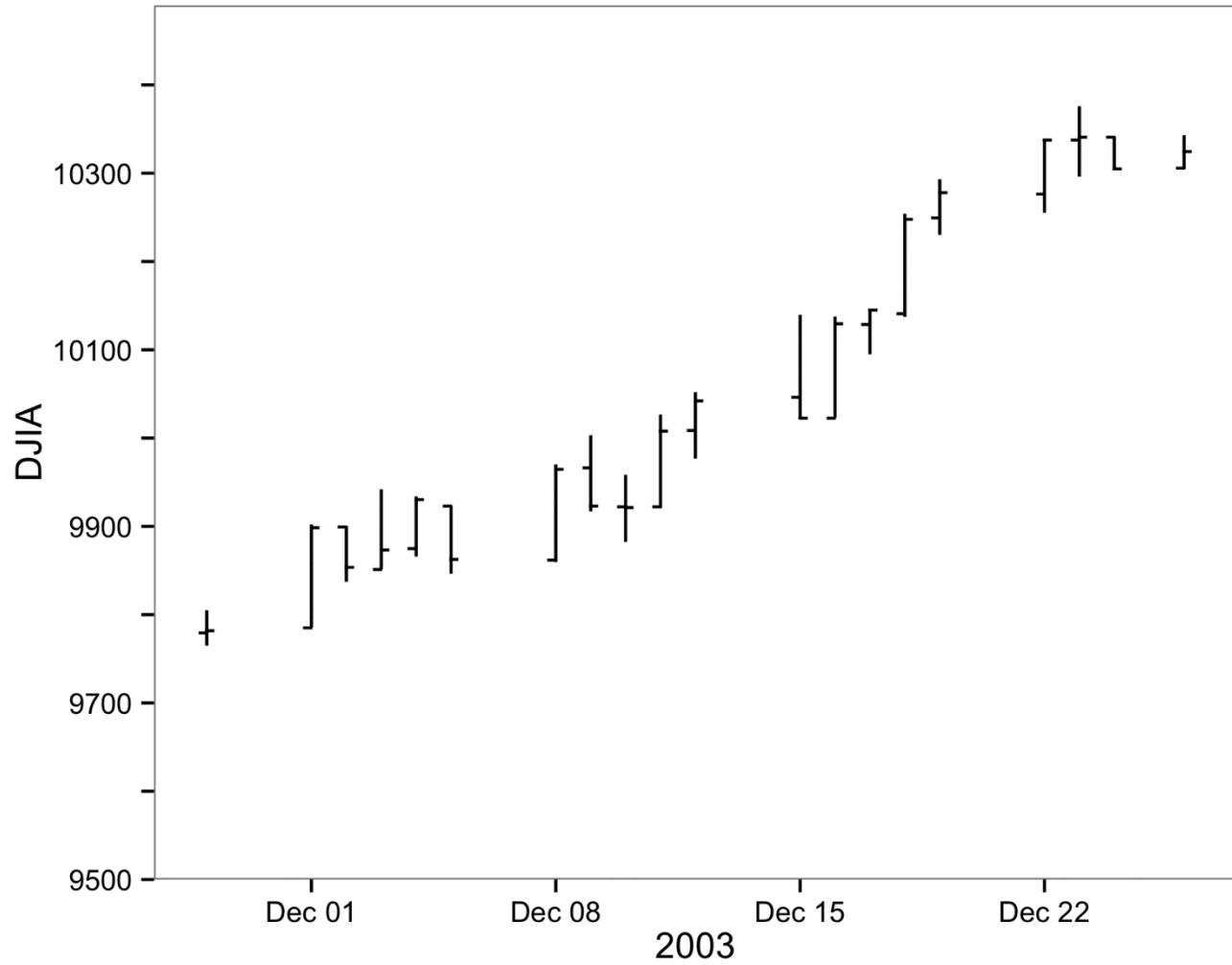
St. Louis Science Center attendance, 1998 to 2002

↑  
quantitative  
variable  
↓



# *financial plot*

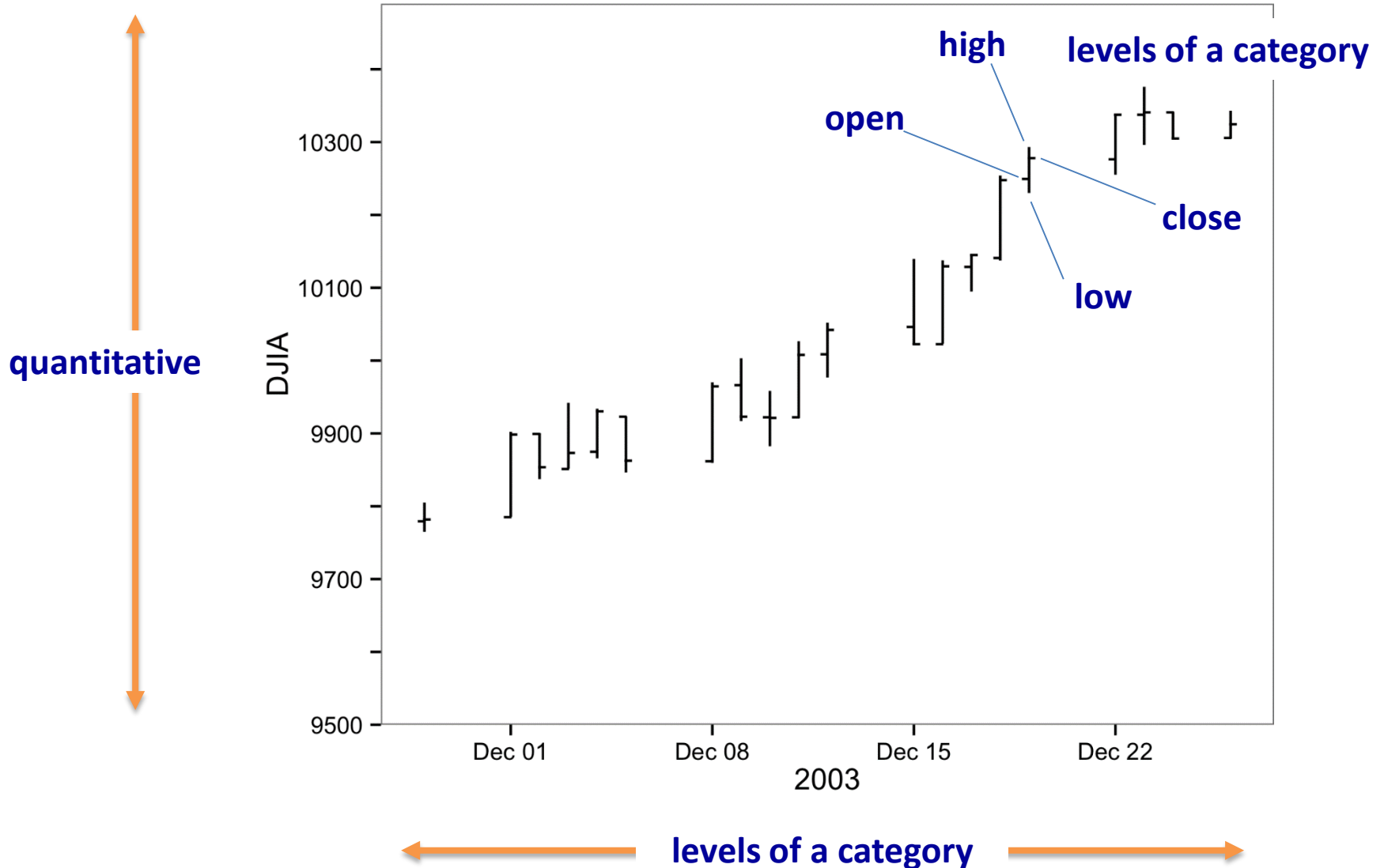
Dow Jones Industrial Average



**Story:** evolution and comparison  
**Data:** 1 quantitative, 2 categorical

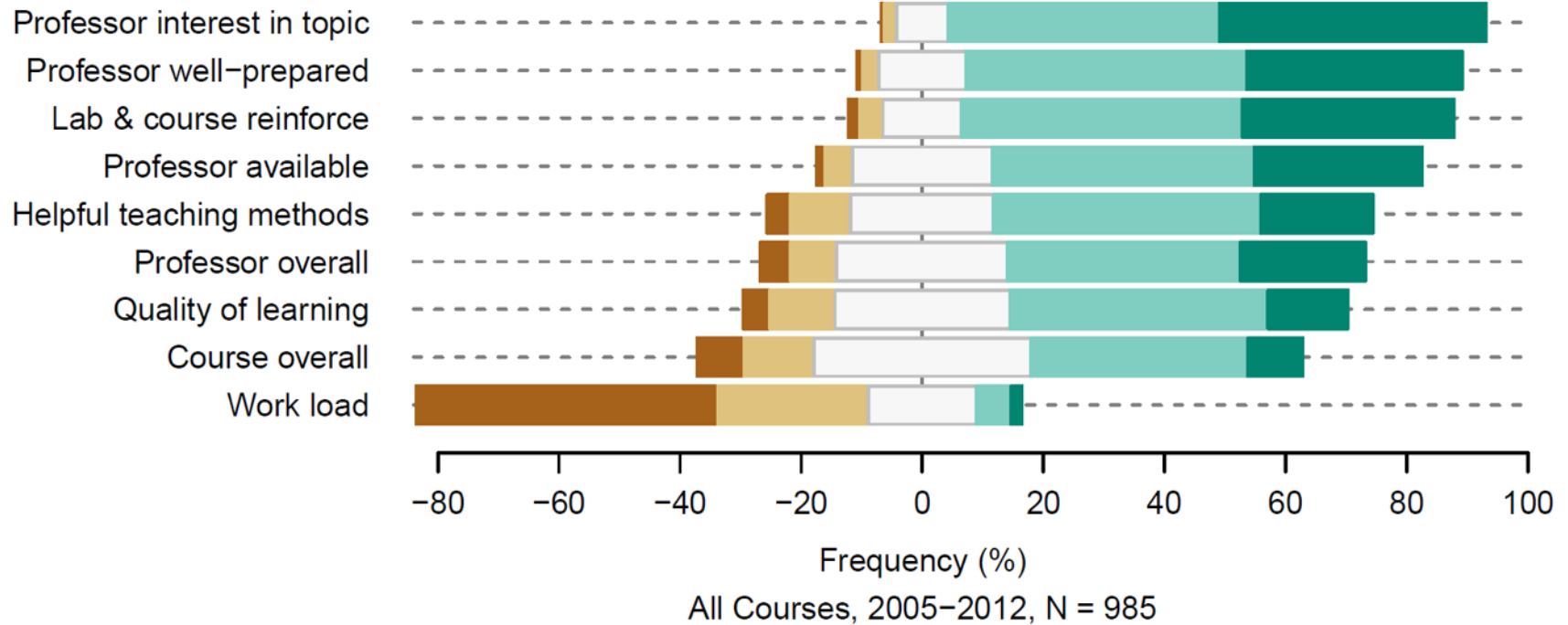
*financial plot*

Dow Jones Industrial Average



# *diverging stacked bar*

Course evaluation survey results

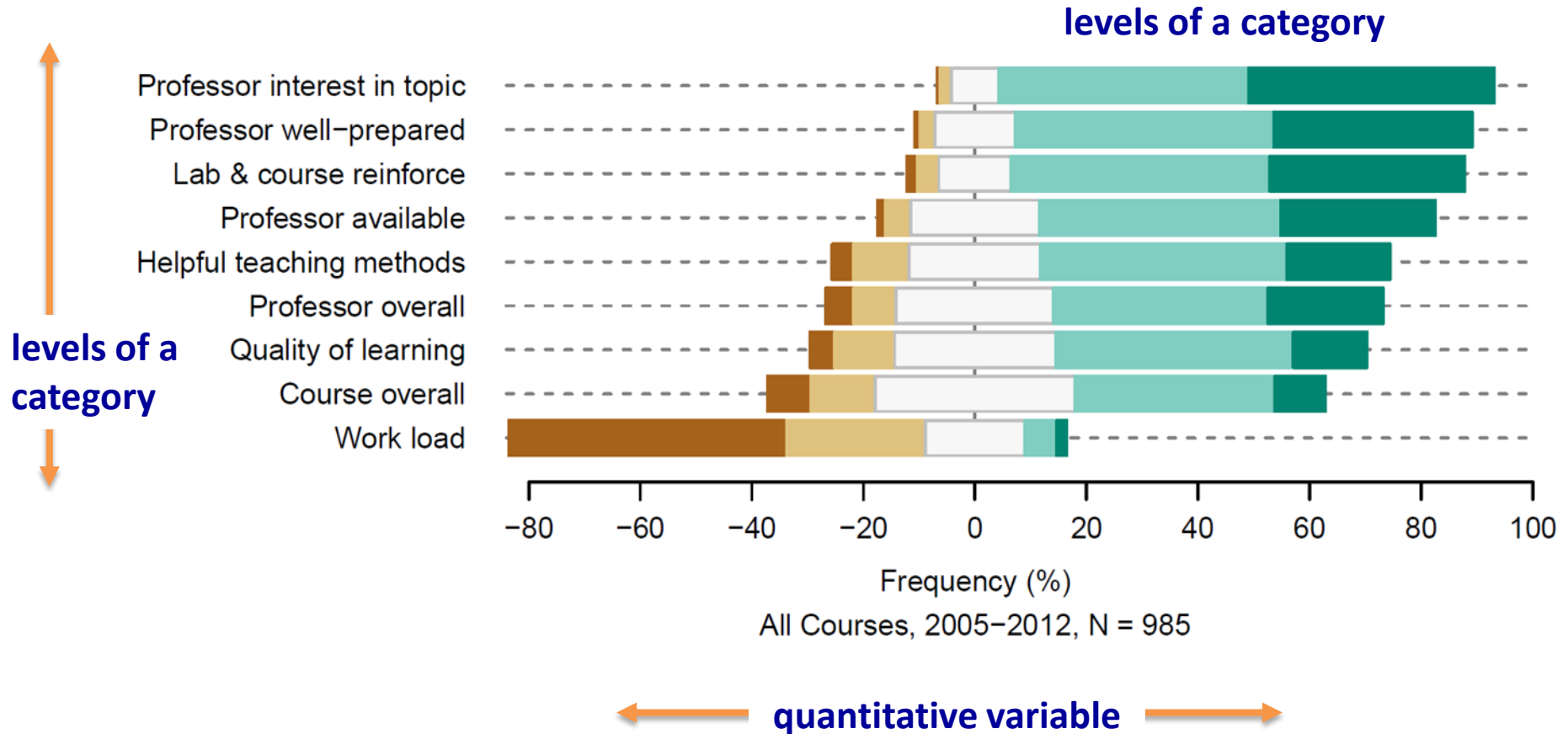


Story: **comparing data**

Data: **1 quantitative, 2 categorical**

# *diverging stacked bar*

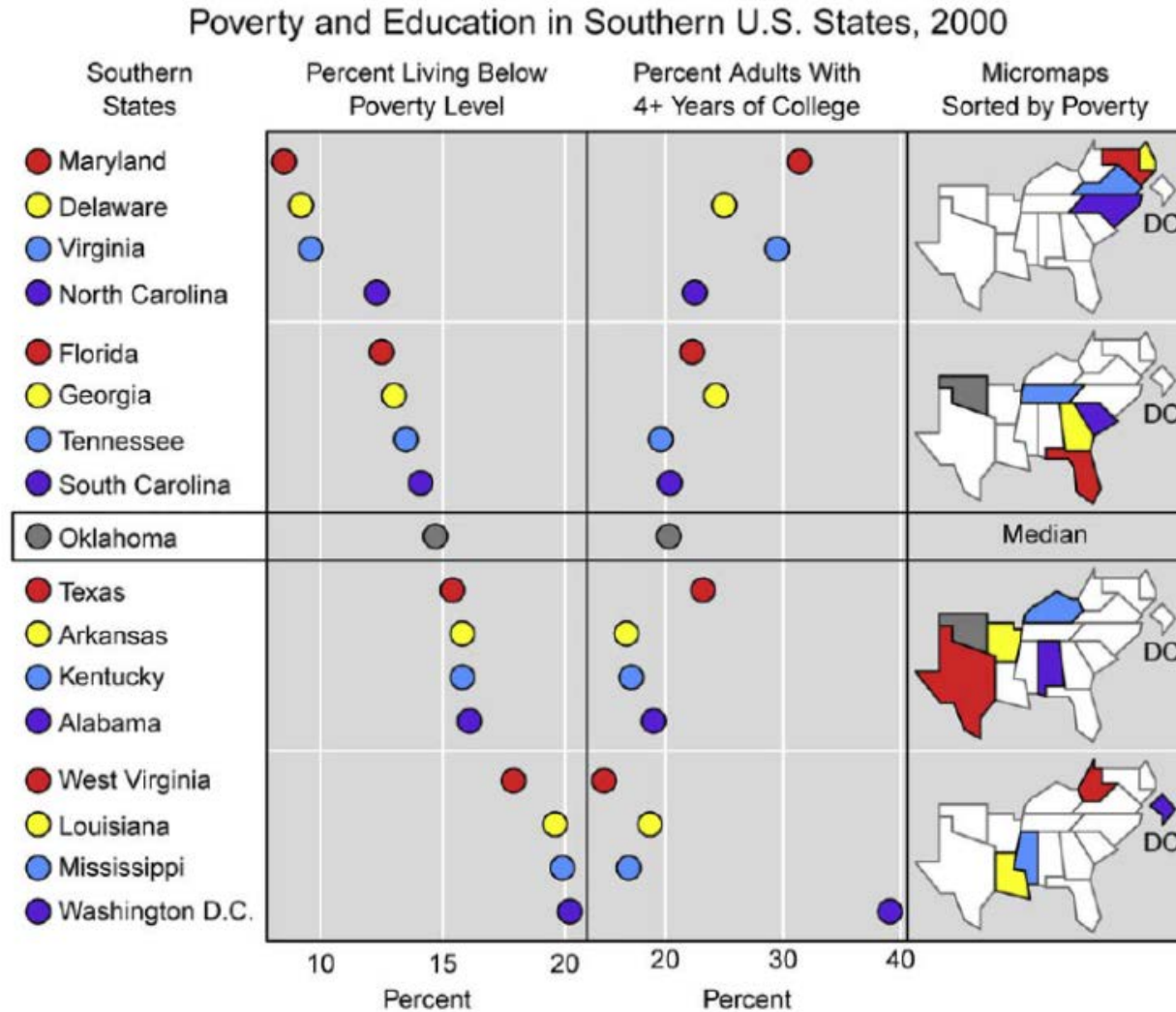
Course evaluation survey results





# linked micromaps

Poverty and education level



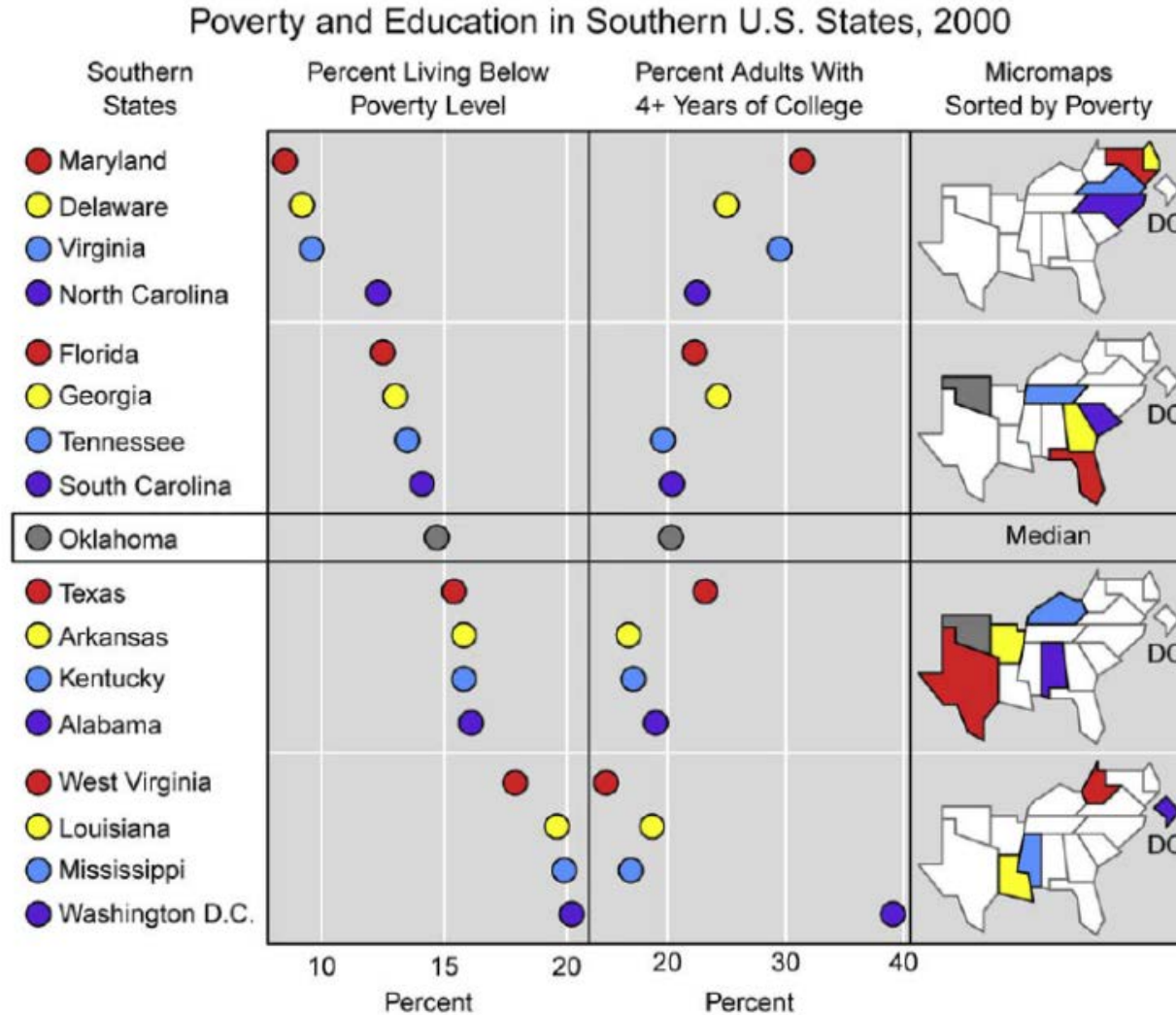
Story: comparing, revealing correlations

Data: 2 quantitative, 2 categorical

# *linked micromaps*

Poverty and education level

↑  
levels of a  
category  
↓



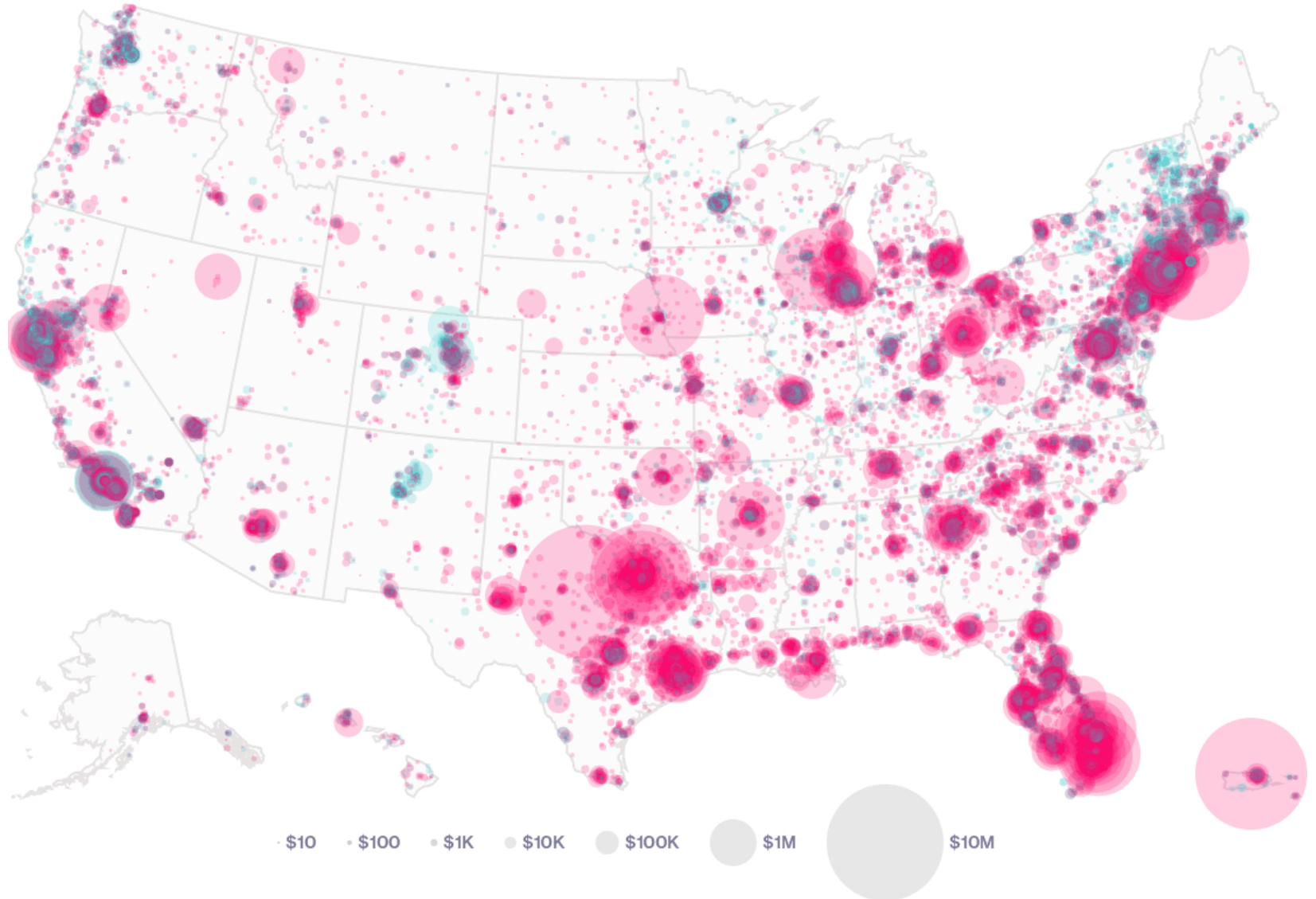
↑  
levels of a  
category  
↓

← quantitative →

← quantitative →

# *proportional symbol*

Presidential election fundraising



**Story:** spatial distribution,  
comparing data

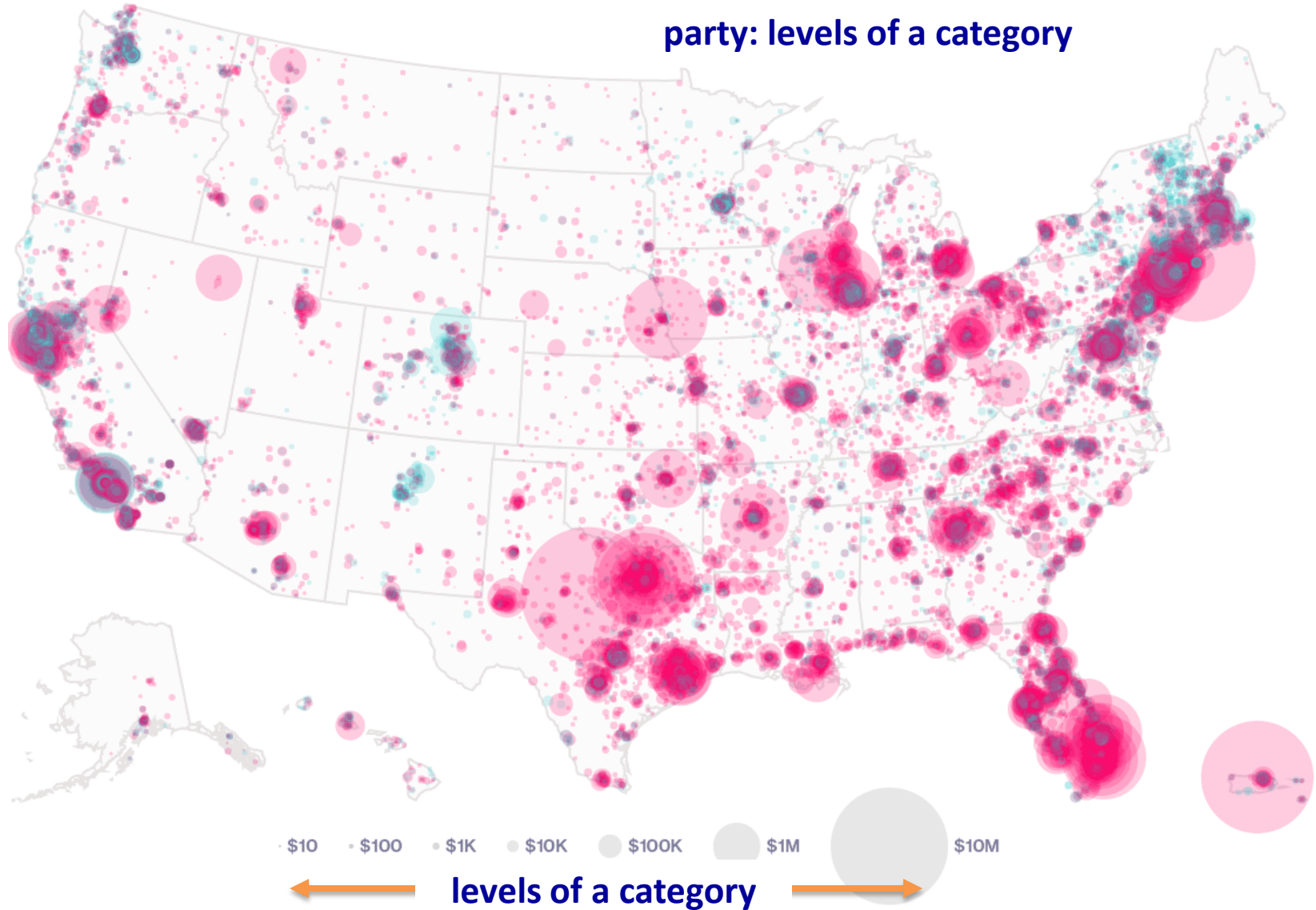
**Data:** 3 categorical

*proportional symbol*

Presidential election fundraising

ZIP code: levels of a category

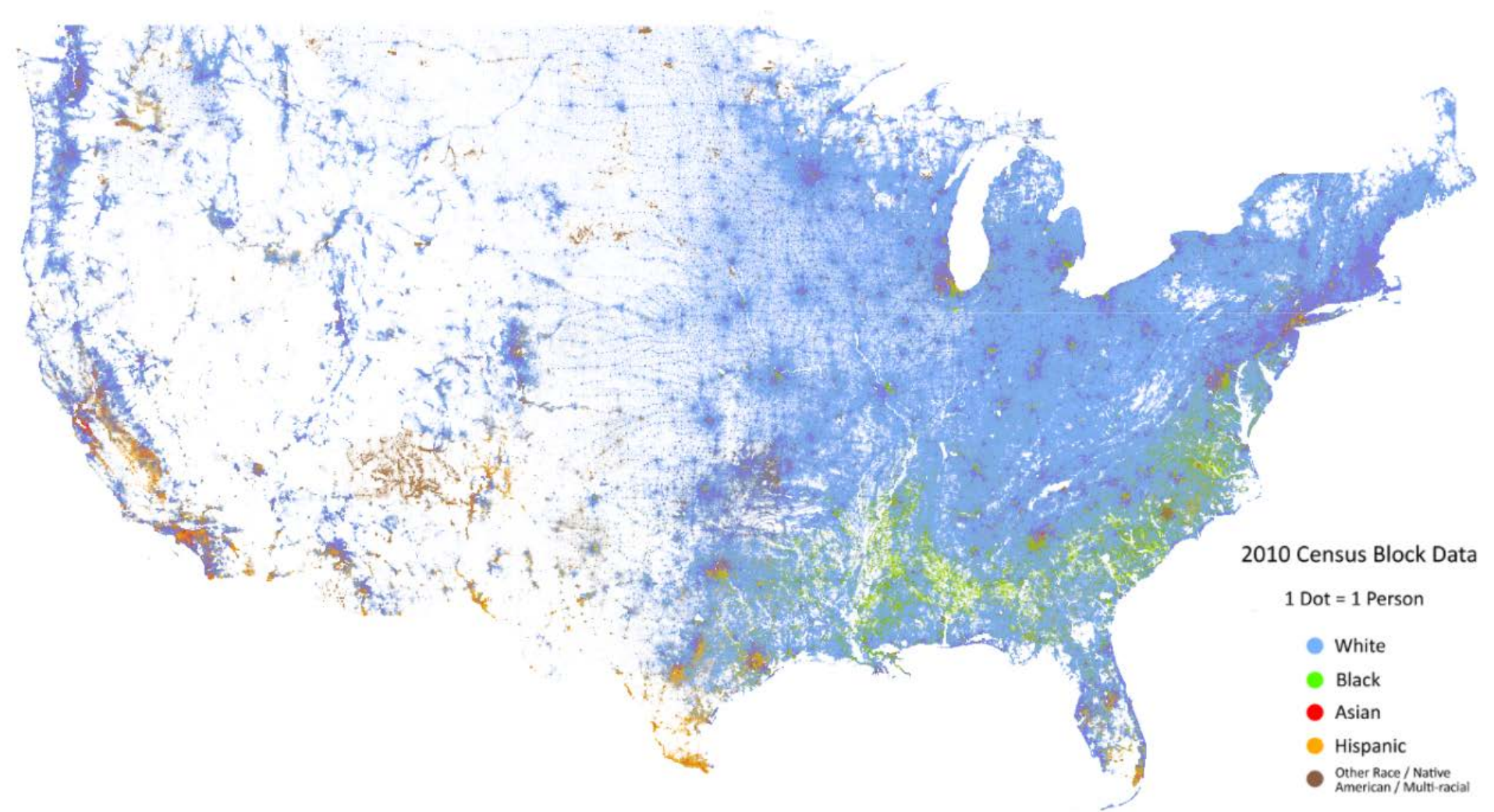
party: levels of a category





# *dot density*

2010 Census population density



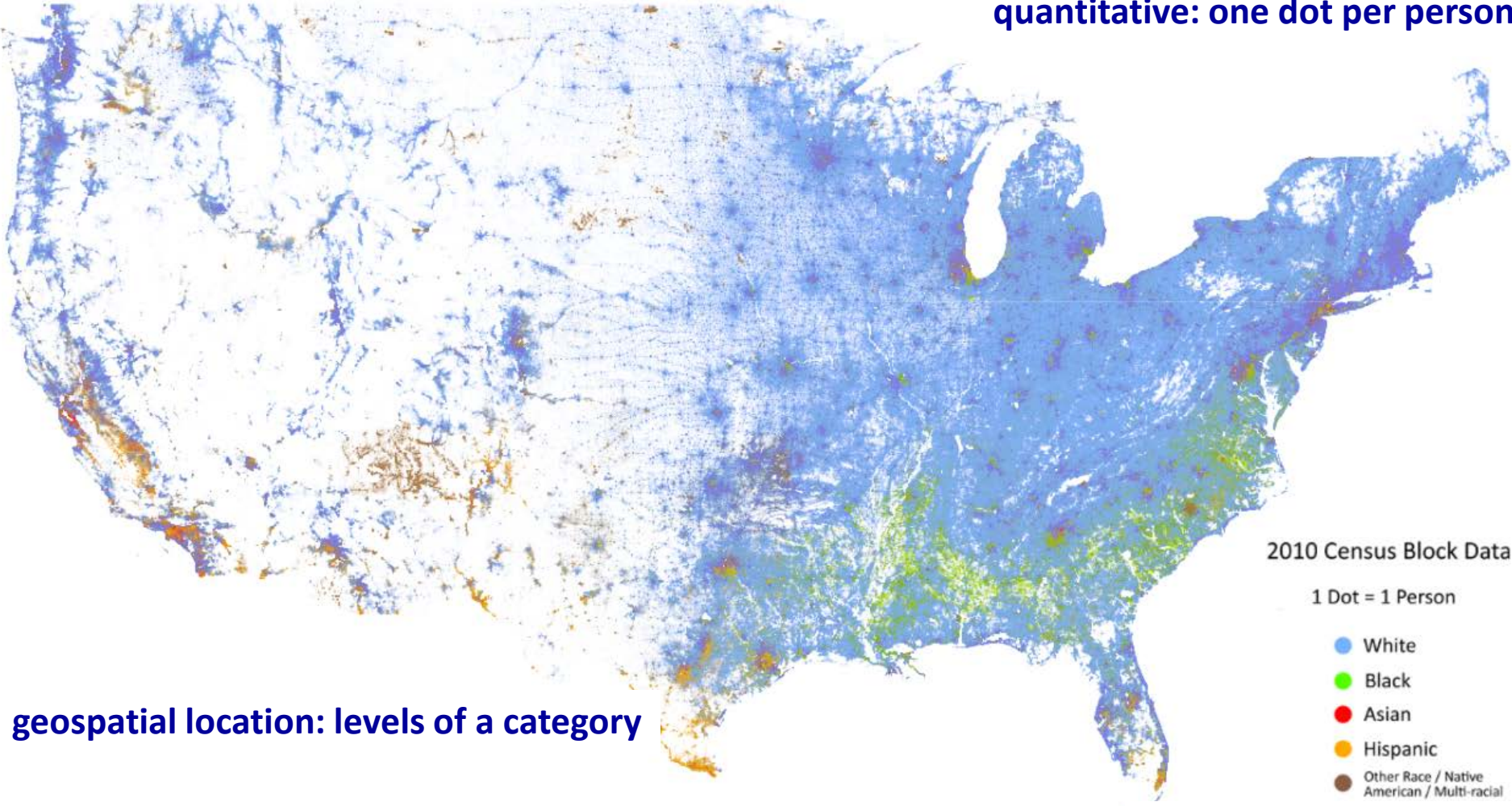
**Story:** spatial distribution, comparing data

**Data:** 1 quantitative, 2 categorical

*dot density*

2010 Census population density

**quantitative: one dot per person**

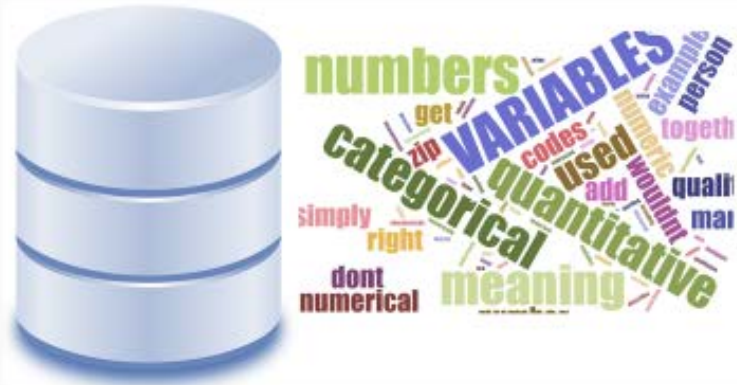


**geospatial location: levels of a category**

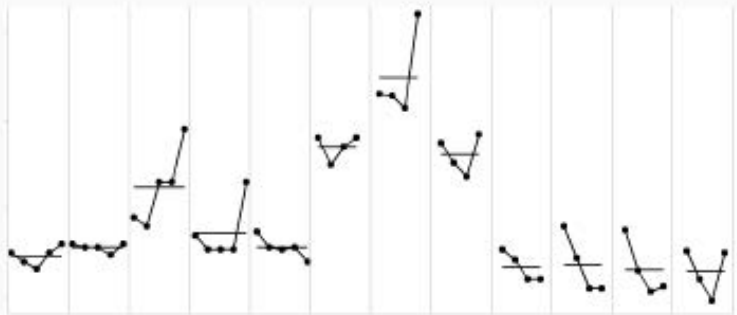
**levels of a category**



# Implications for the designer



Grasp the **data structure** first



Explore **data-suitable** designs