

---

# BUSINESS INTELLIGENCE



---

## LECTURE 4

---

# OUTLINE

- Introduction
- Business Intelligence Systems
- Fundamental of Visualization
- References

# BI FEATURES

- BI task
  - Providing decision support for specific goals defined in the context of business activities in different domain areas.
- BI foundation
  - BI decision support mainly relies on empirical information based on data.
- BI realization
  - The decision support has to be realized as a system using the actual capabilities in information and communication technologies (ICT).
- BI delivery
  - A BI system has to deliver information at the right time to the right people in an appropriate form.

# BI CONTEXT

## ■ By scenarios

- ❑ Standardized reports for a dedicated part of the business
- ❑ Performance monitoring
- ❑ A feedback on strategy formulation
- ❑ Strategic resource

## ■ By perspective

- ❑ Production
- ❑ Customer
- ❑ Organization
- ❑ Etc.

---

# GOALS OF BI

- Key Performance Indicator (KPI)
    - A KPI links the activities of the business to objectives by defining a measurable quantity.
  - Influential factors
    - Attributes that may influence the behavior of the KPI in any BI perspective.
  - Analytical goals
    - Descriptive: reporting, segmentation, interesting behavior detection
    - Predictive: regression, classification
-

# DISCUSSION



## BI TASKS TO DO

---

# WHAT WOULD WE DO?

- Sales dashboard
  - Sales forecast
- Delivery performance
- Reservation performance
- Customer dashboard
  - Repeat rate
- Feedback dashboard
  - Feedback score
- Item dashboard
  - Item ABC analysis
- Etc.

---

# BI TASKS

- Data task
- Business and data understanding task
- Modeling task
- Analysis task
- Evaluation and reporting task



---

# DATA TASK

- A prerequisite for all BI activities
- The main goal is to organize available information about the business and its environment
- Not only retrieve data but also reorganize and collect additional data for special purposes

---

# BUSINESS AND DATA UNDERSTANDING TASK

- Determine BI context in business
- An initial formulation of goals and KPIs
- Data needed to achieve those goals

# MODELING TASK

- It aims at setting up an analytical business model, i.e., a formal model that allows precise answers for the analytical goals, and data related models

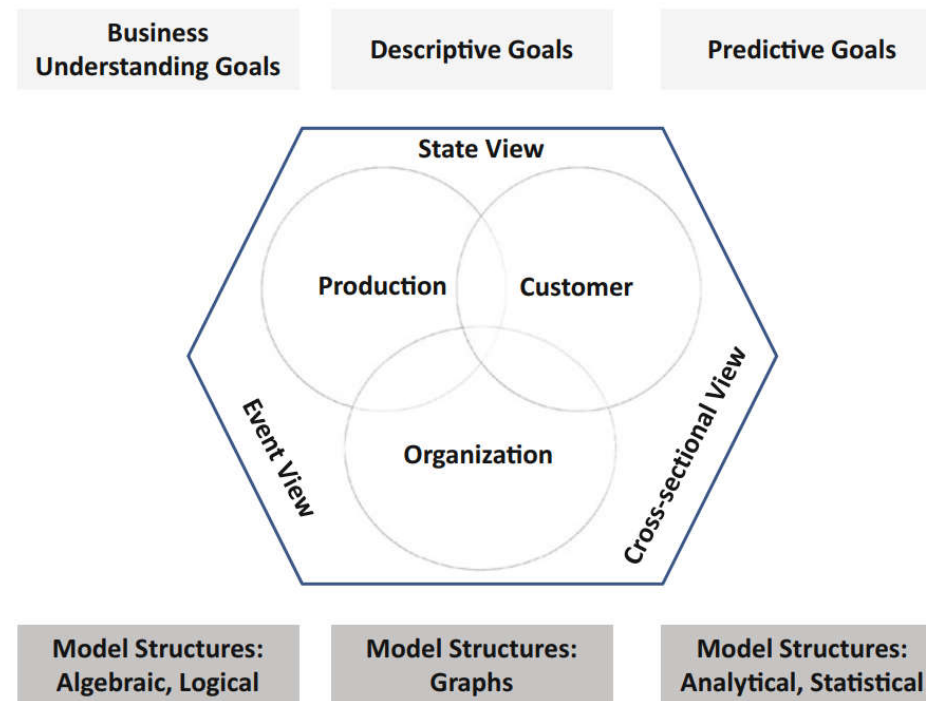


Fig. 1.2 Overview on modeling activities

---

# ANALYSIS TASK

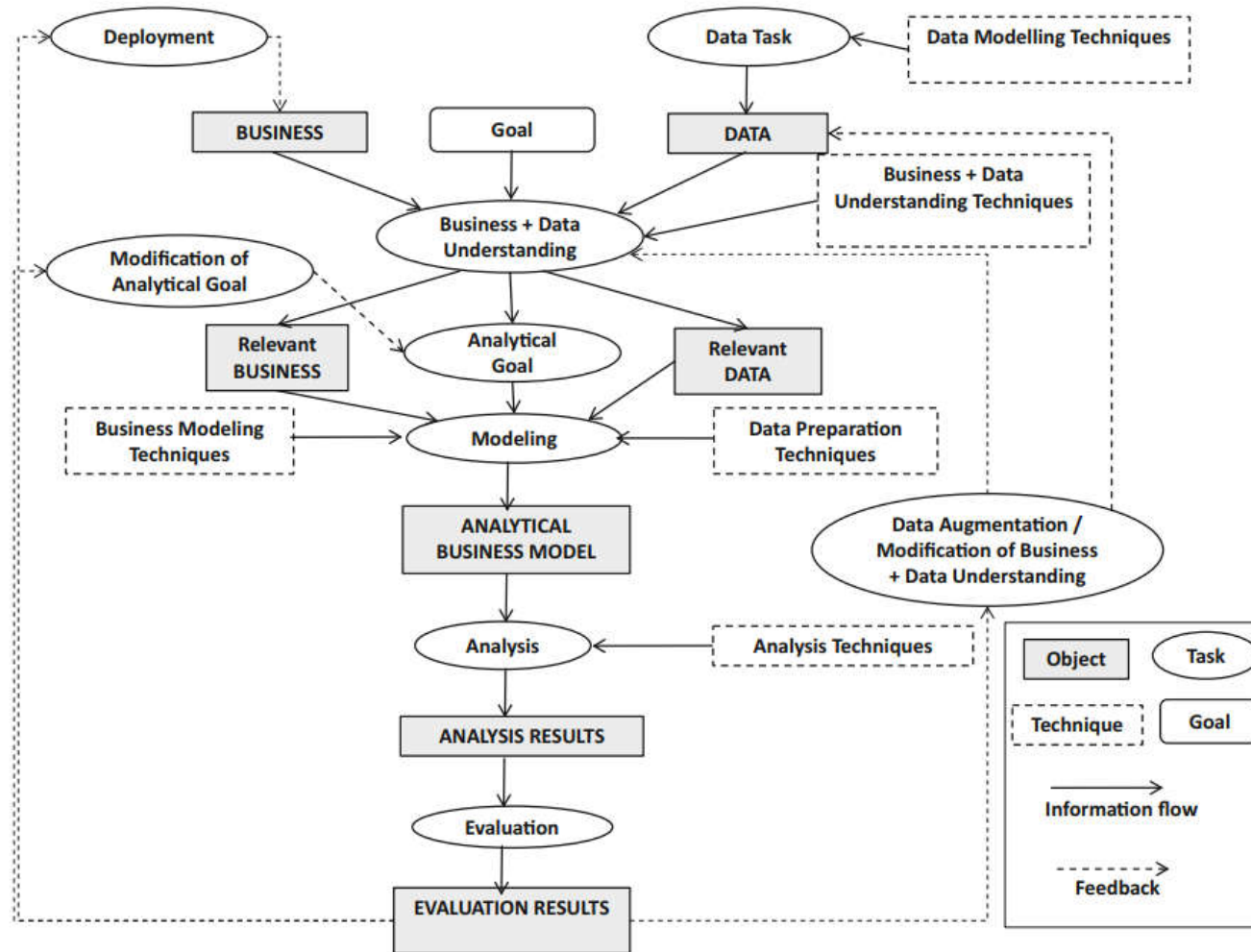
- Algorithms to compute a solution for the analytical goal within the model.
  - Statistics methods
  - Data mining
  - Machine learning
  - Etc.

---

# EVALUATION AND REPORTING TASK

- Evaluation task: view the analysis results from two different perspectives
  - The context of analytical goal
  - The context of business
- Reporting task: interpret the results to domain knowledge by description and visualization

# THE iMine METHOD



**Fig. 1.4** The *iMine* method

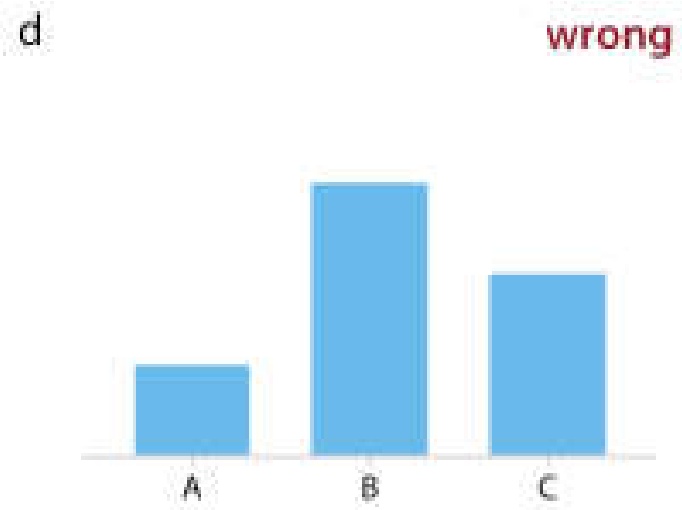
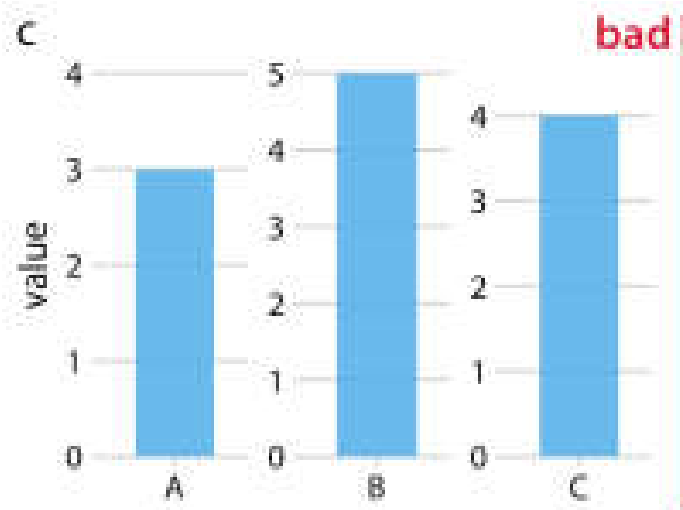
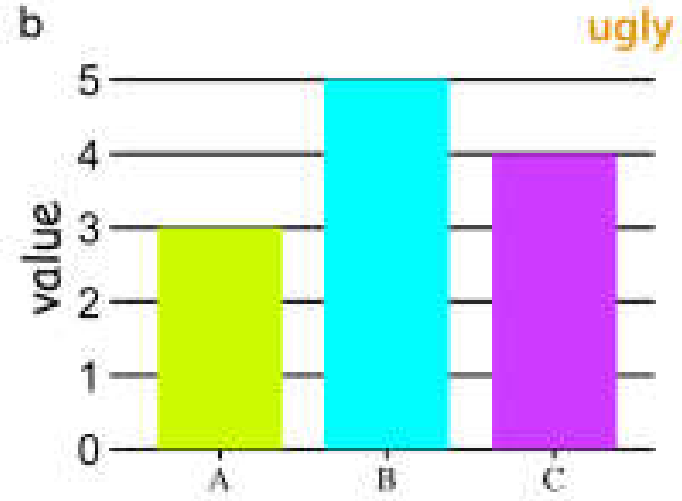
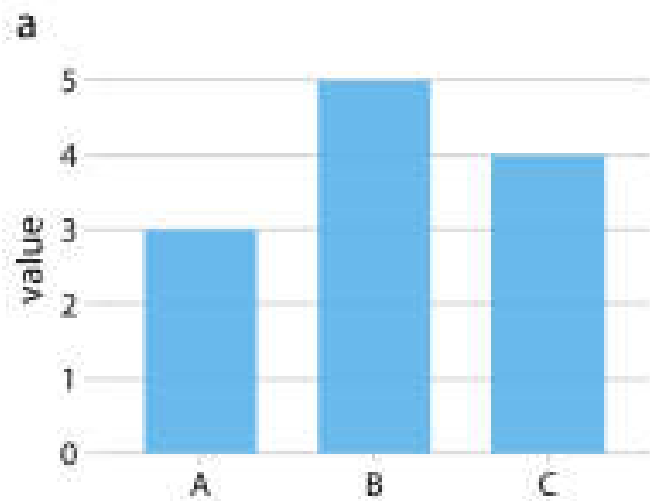
---

# DISCUSSION



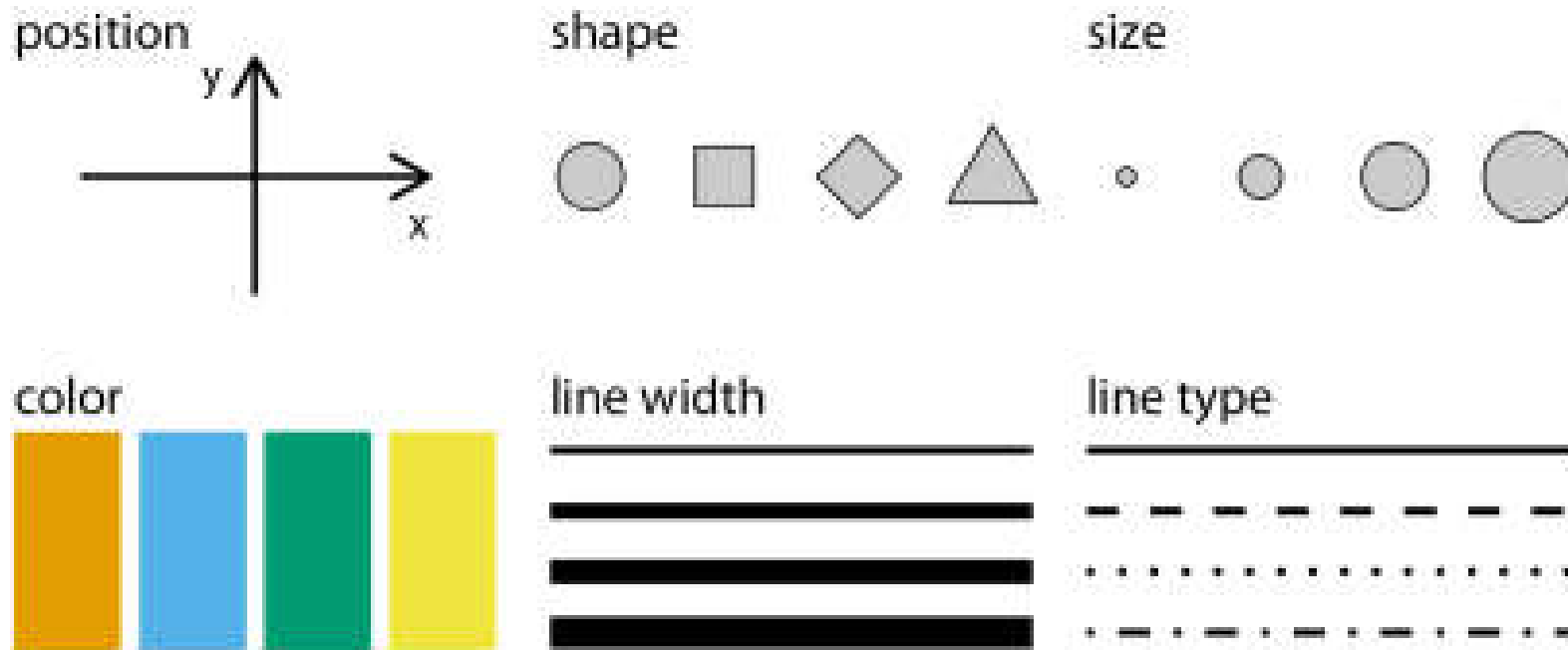
## DELIVERY DASHBOARD

# CHART PROBLEMS





# AESTHETICS IN DATA VISUALIZATION



# TYPES OF DATA

Type of variable	Appropriate scale	Example
Quantitative, numerical continuous	Continuous	1.2, 5.8, 10, 3e-2
Quantitative, numerical discrete	Discrete	1, 5, 8, 12
Qualitative, categorical unordered	Discrete	Rose, Violet, Blossom
Qualitative, categorical ordered	Discrete	Bad, good, excellent
Datetime	Discrete or Continuous	2023-02-23 04:35:58
Text	None or Discrete	The cake is too burnt

# DISCUSSION



## DATA VISUALIZATION

Month	Day	Location	Station ID	Temperature (°F)
Jan	1	Chicago	USW00014819	25.6
Jan	1	San Diego	USW00093107	55.2
Jan	1	Houston	USW00012918	53.9
Jan	1	Death Valley	USC00042319	51.0
Jan	2	Chicago	USW00014819	25.5
Jan	2	San Diego	USW00093107	55.3
Jan	2	Houston	USW00012918	53.8
Jan	2	Death Valley	USC00042319	51.2

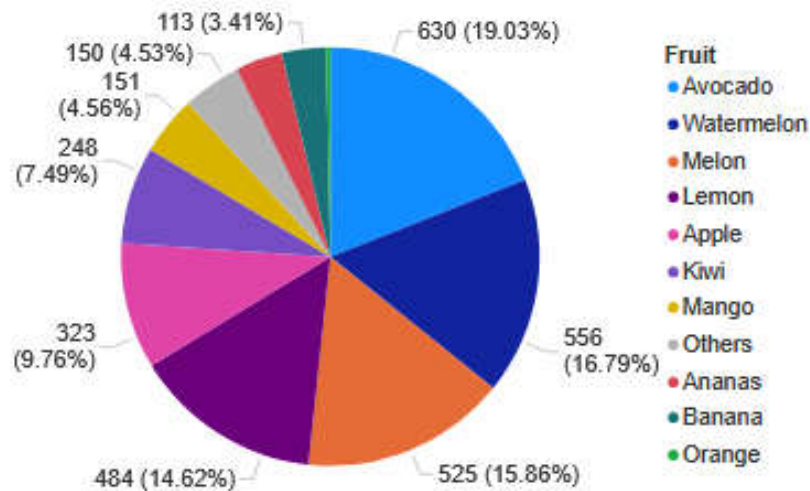
Rank	Title	Weekend gross
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746

Country	Year	Status	Life expectancy
Belgium	2001	Developed	78
Lithuania	2000	Developed	71.6
Iceland	2013	Developed	82.4
Iceland	2012	Developed	82.5
Australia	2014	Developed	82.7
Cyprus	2009	Developed	79.3
Italy	2012	Developed	82
Italy	2011	Developed	82
Italy	2010	Developed	81.8
Netherlands	2013	Developed	81.4
Singapore	2011	Developed	82.2
Spain	2013	Developed	82.4
Switzerland	2009	Developed	82.1
Switzerland	2008	Developed	82
Austria	2012	Developed	88
Cyprus	2000	Developed	78.1
Iceland	2002	Developed	84
Ireland	2011	Developed	84
Japan	2003	Developed	81.9
Japan	2002	Developed	81.8
Luxembourg	2012	Developed	81.1
New Zealand	2012	Developed	81.1
Norway	2006	Developed	84
Spain	2008	Developed	81.3
Sweden	2004	Developed	83
Austria	2007	Developed	81
Belgium	2011	Developed	83
Germany	2008	Developed	79.9
Luxembourg	2008	Developed	80

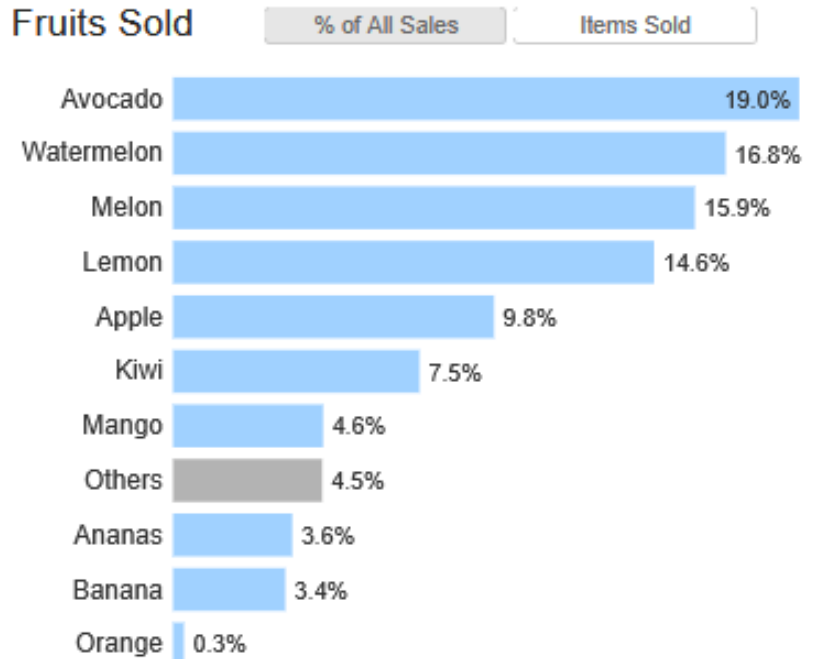
Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# BAD VS. GOOD CHARTS

Fruits Sold

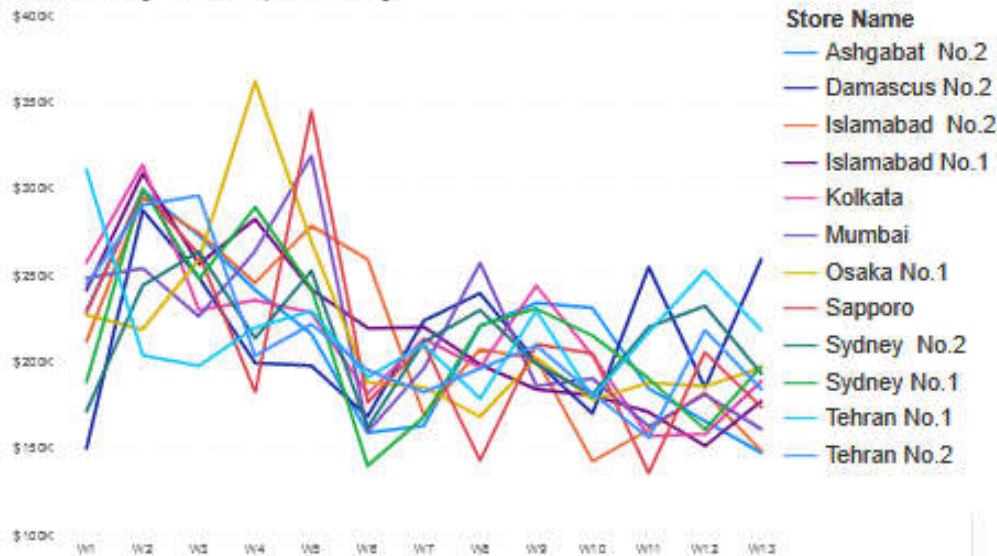


Fruits Sold



# BAD VS. GOOD CHARTS

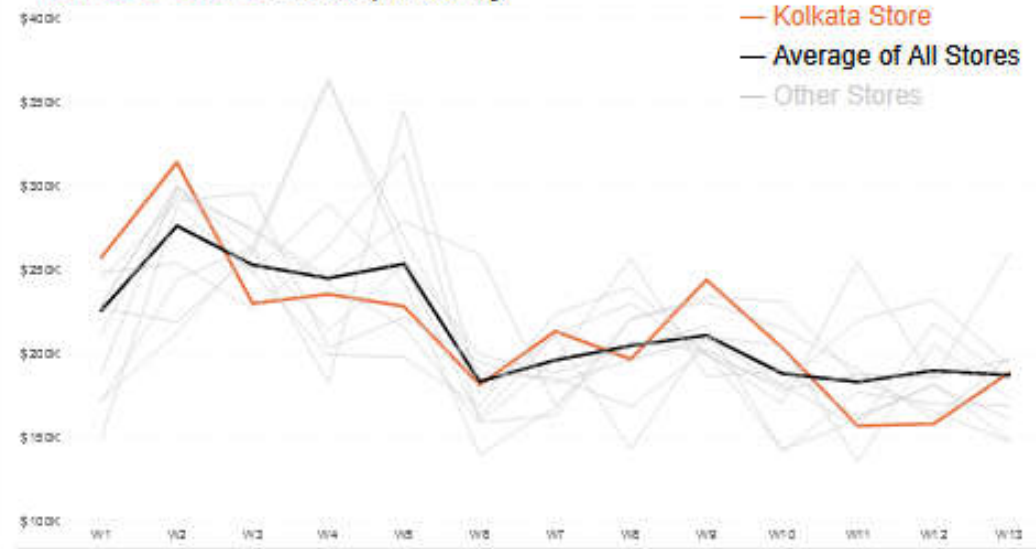
Sales by Store, weekly



Select a store:

Kolkata

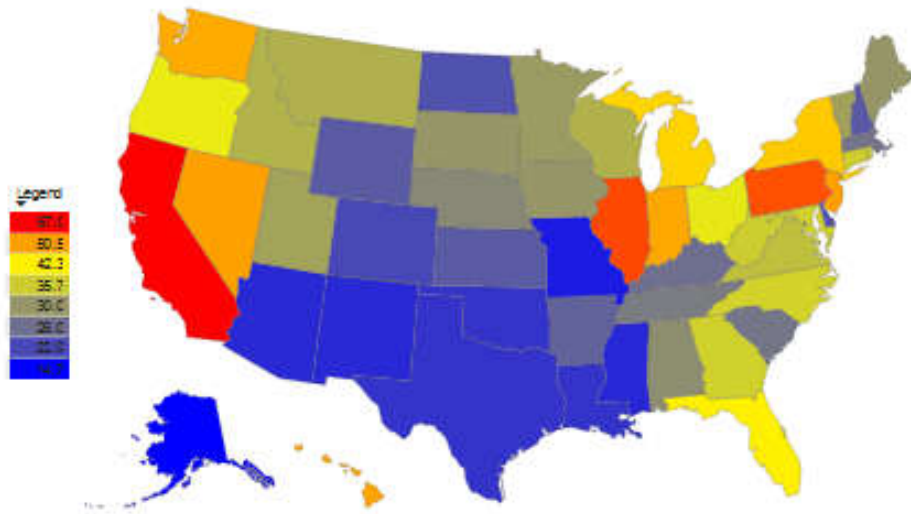
Kolkata Store Sales, weekly



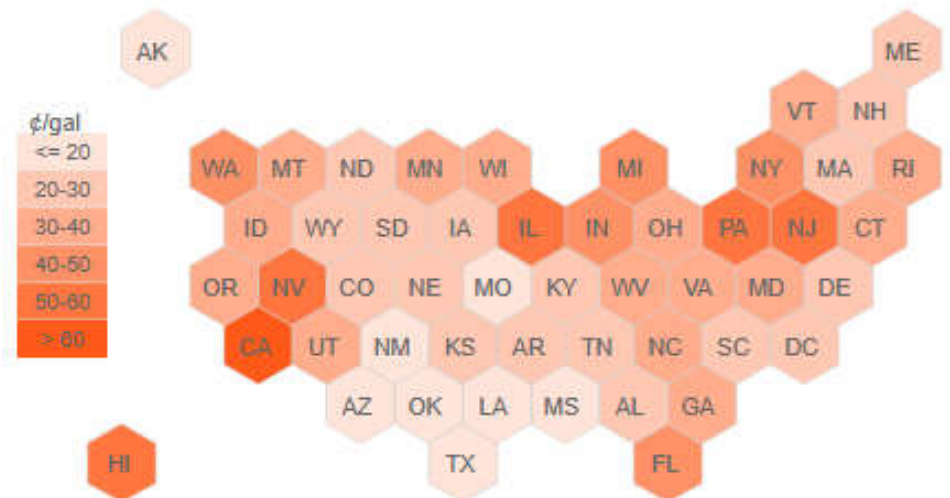
<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

# BAD VS. GOOD CHARTS

Gasoline Tax by US State as of July 2021



Gasoline Tax by US State as of July 2021

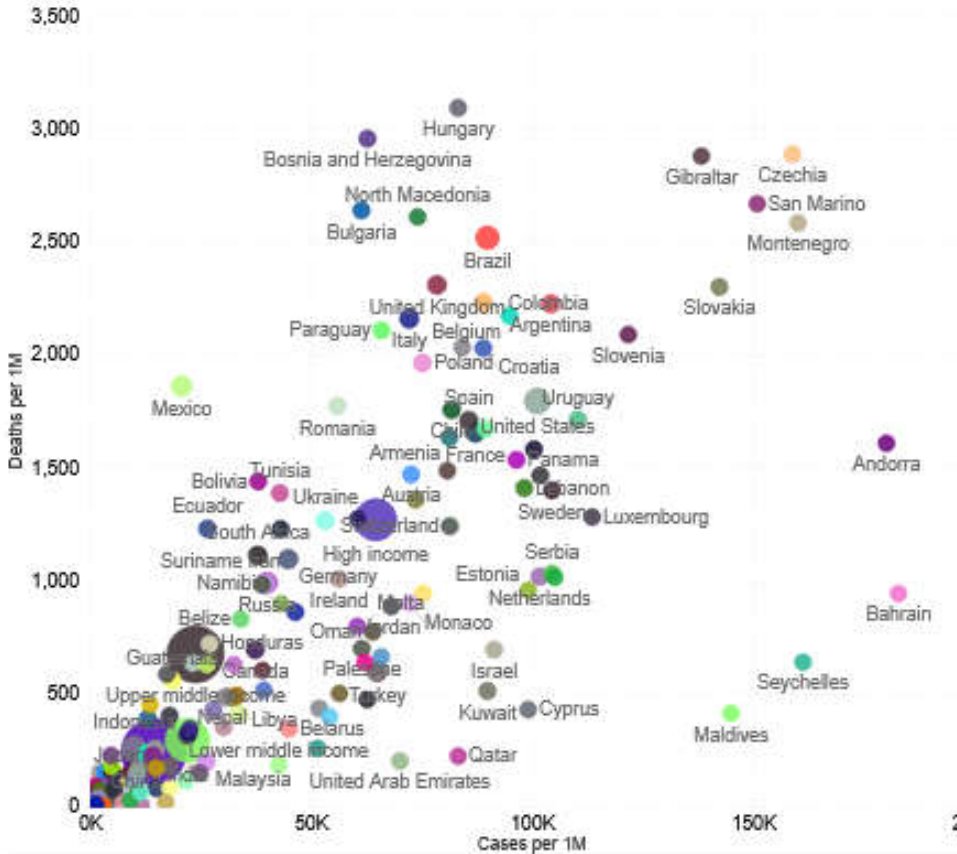


<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

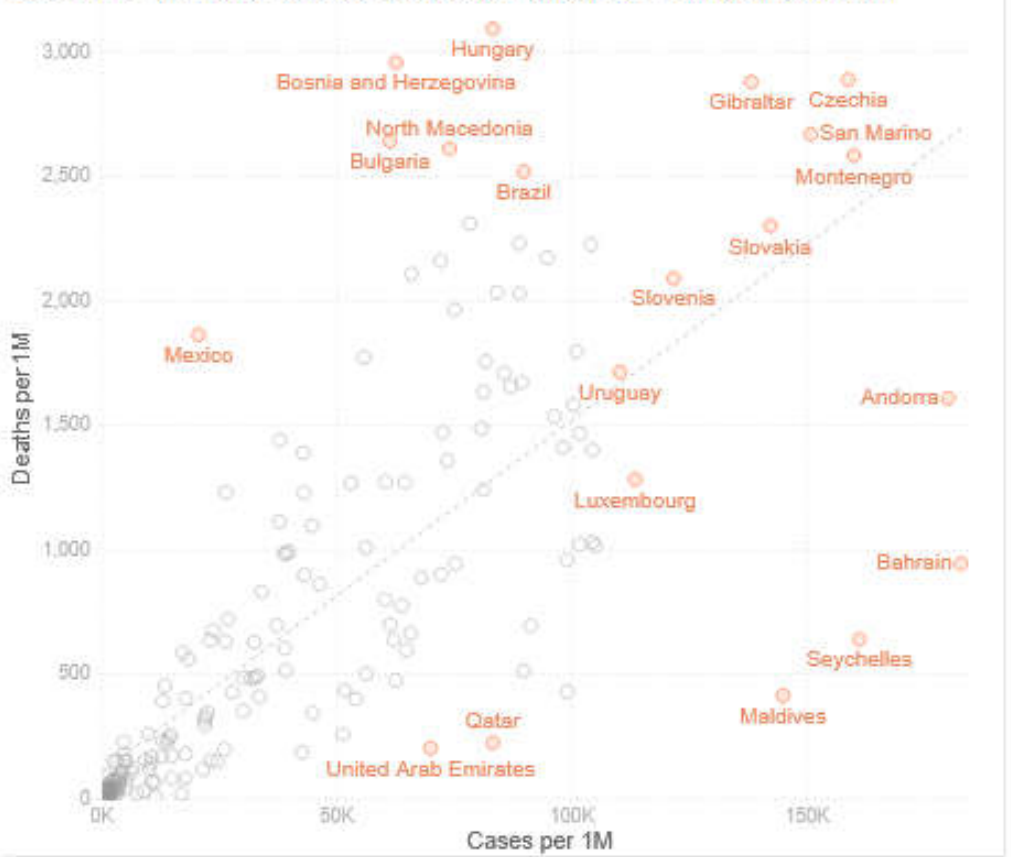


# BAD VS. GOOD CHARTS

COVID-19 total cases and deaths per 1M population



COVID-19 total cases and deaths per 1M population



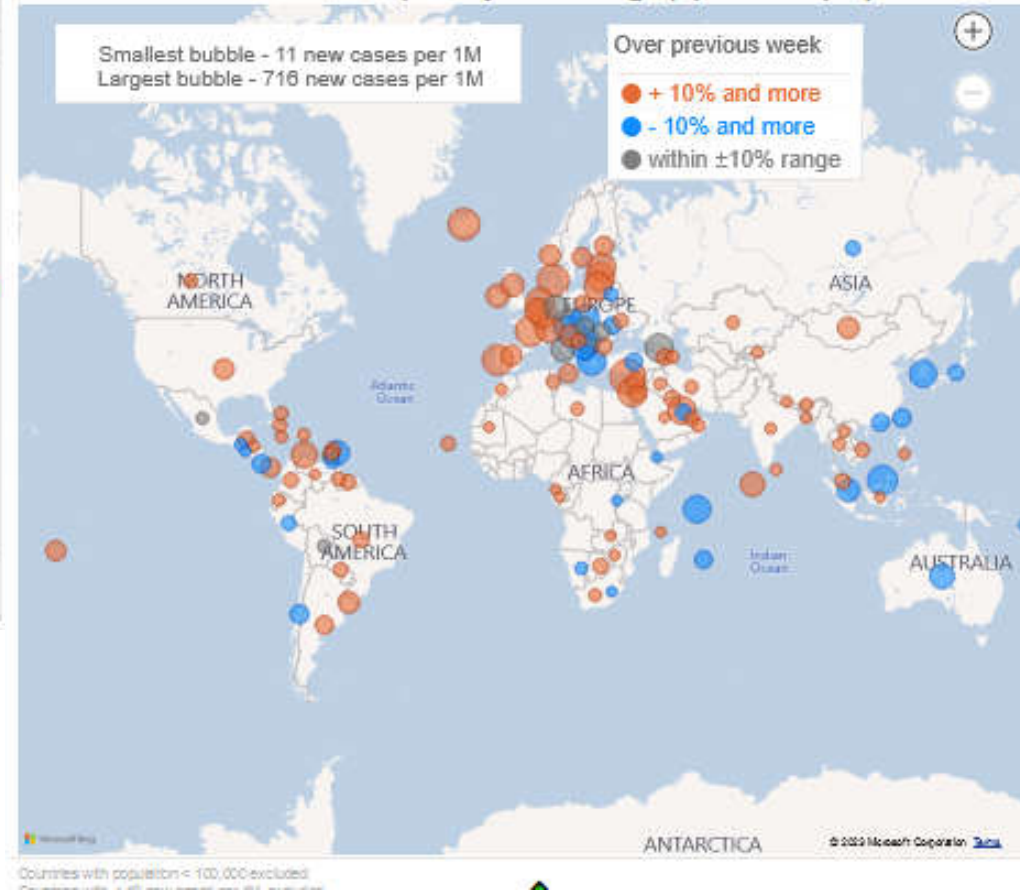
<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

# BAD VS. GOOD CHARTS

COVID-19 total cases per 1M population



COVID-19 new cases (7 days average) per 1M population



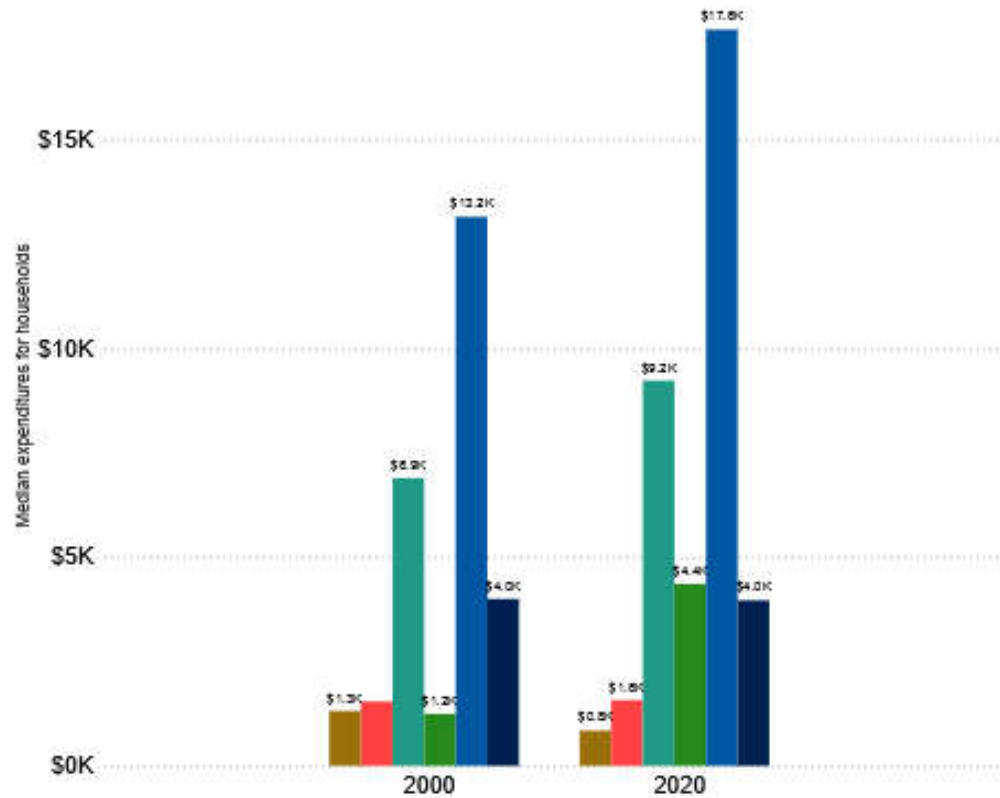
<https://powerofbi.org/bad-and-good-charts-in-power-bi/>



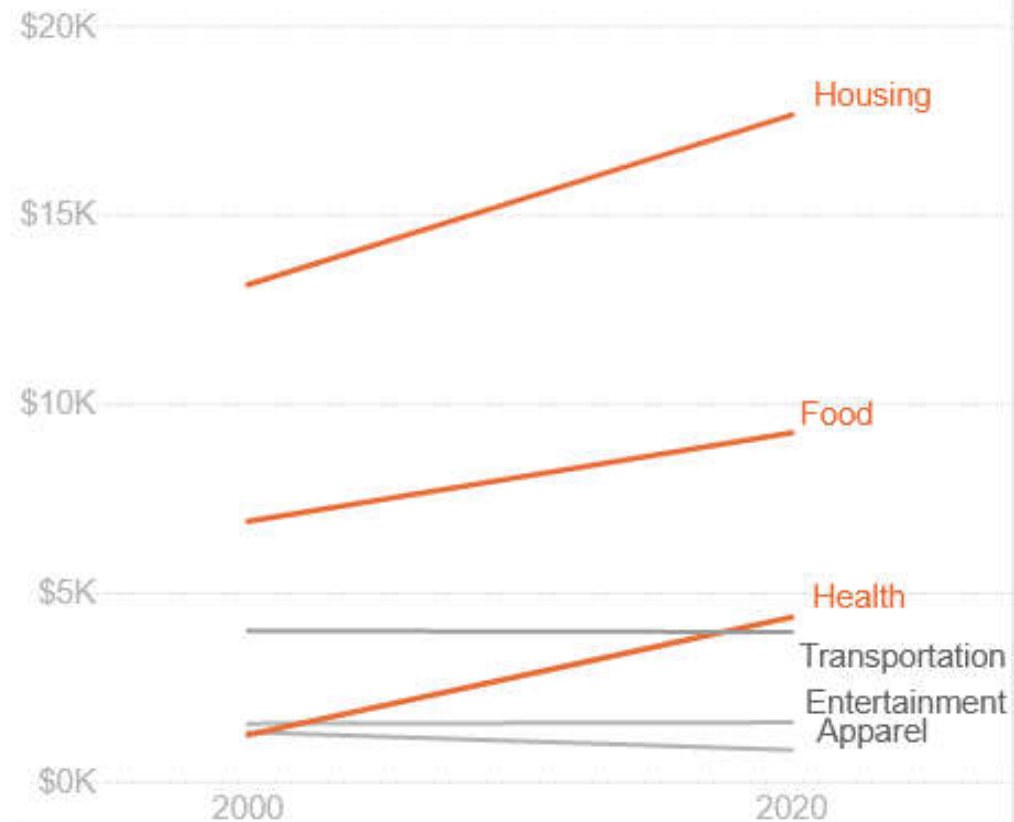
# BAD VS. GOOD CHARTS

Family Expenses, 2020 vs 2000

Category • Apparel • Entertainment • Food • Health • Housing • Transportation



Family Expenses, 2020 vs 2000

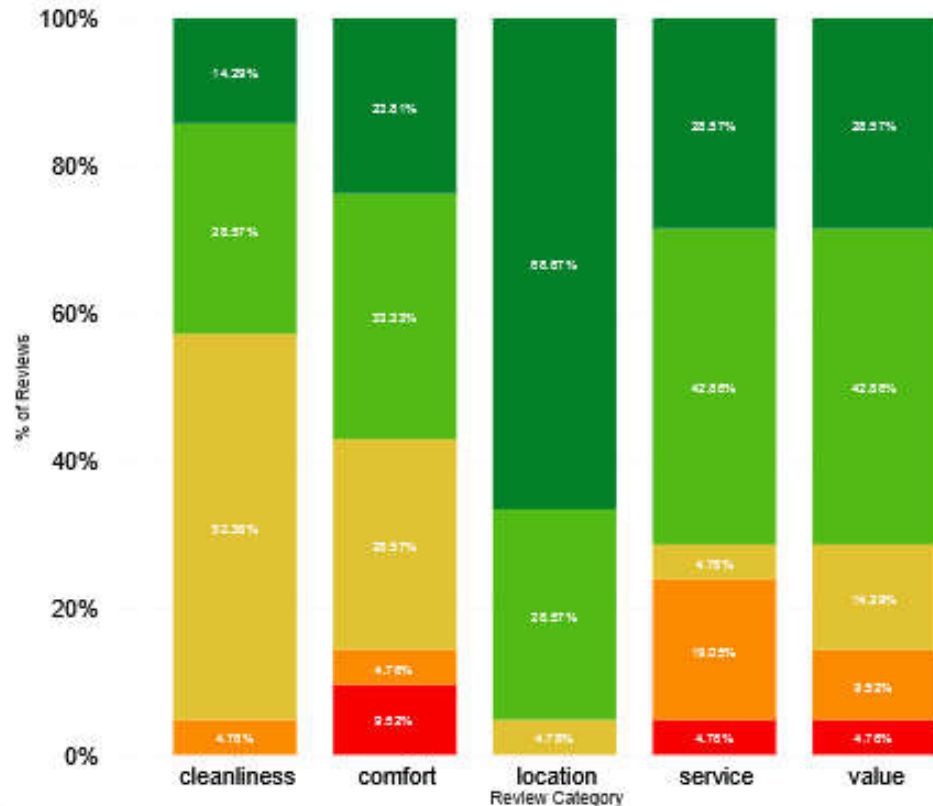


<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

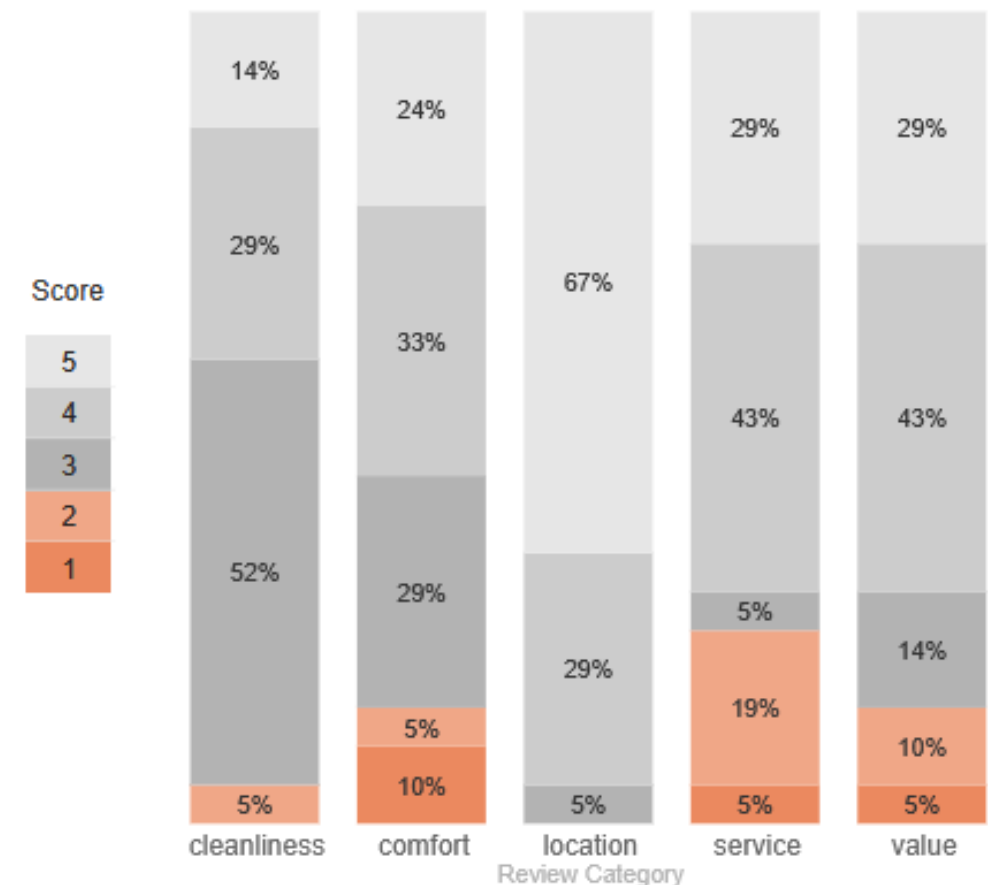
# BAD VS. GOOD CHARTS

Customer Reviews

Score ● 1 ● 2 ● 3 ● 4 ● 5

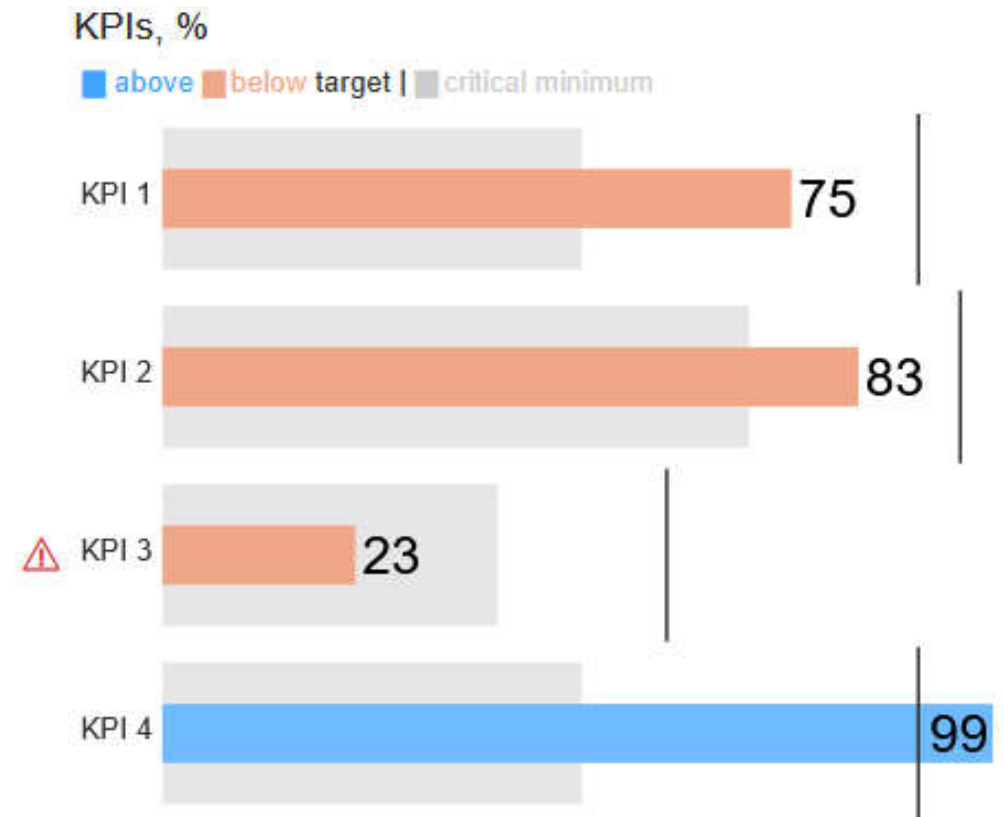
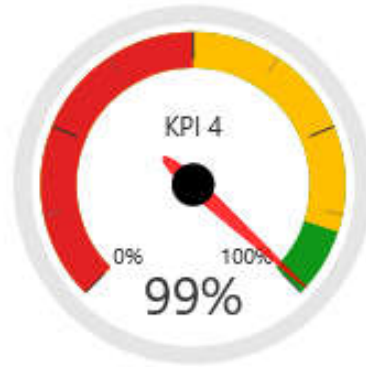
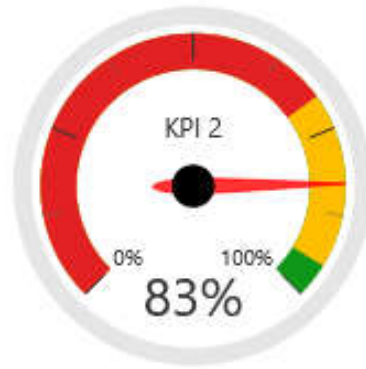
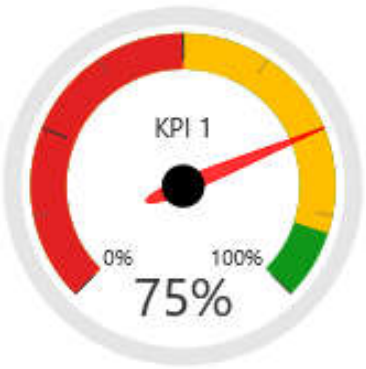


Bad Customer Reviews

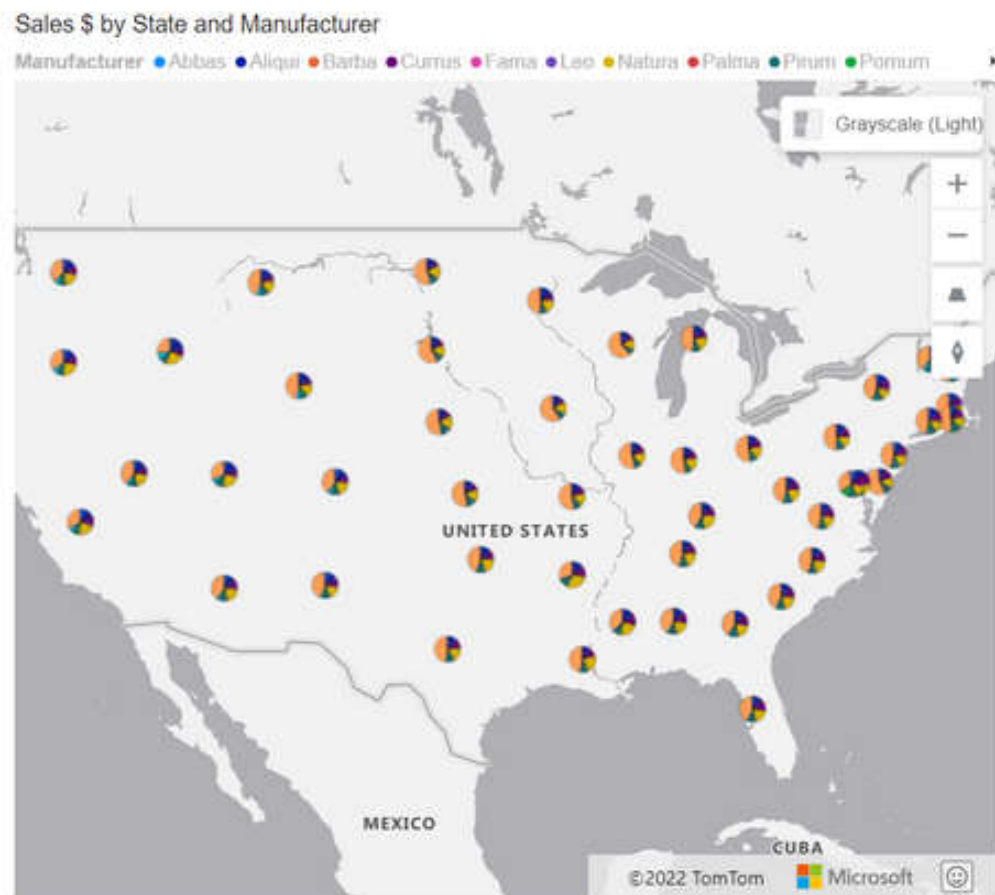


<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

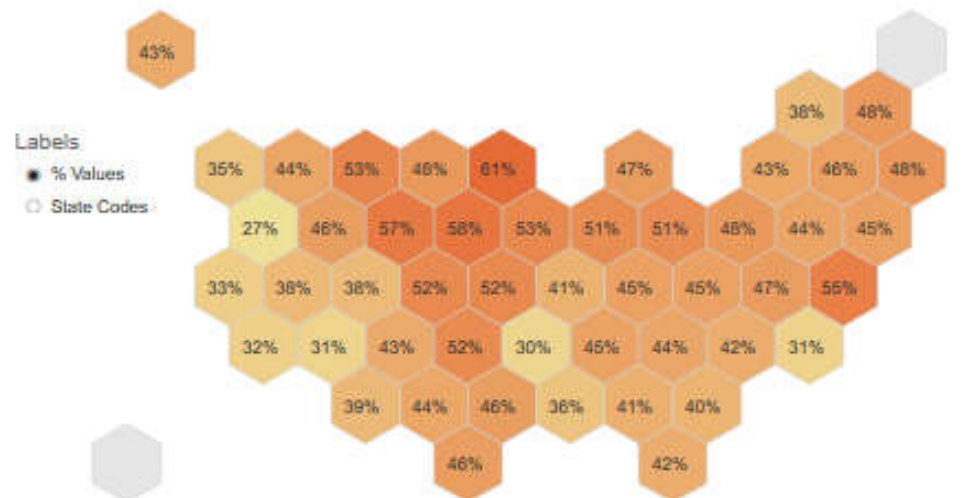
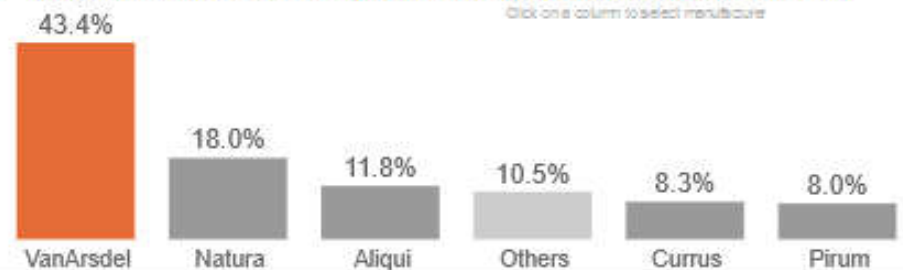
# BAD VS. GOOD CHARTS



# BAD VS. GOOD CHARTS



US Total Market Share (revenue) by Manufacturer, %



<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

# BAD VS. GOOD CHARTS

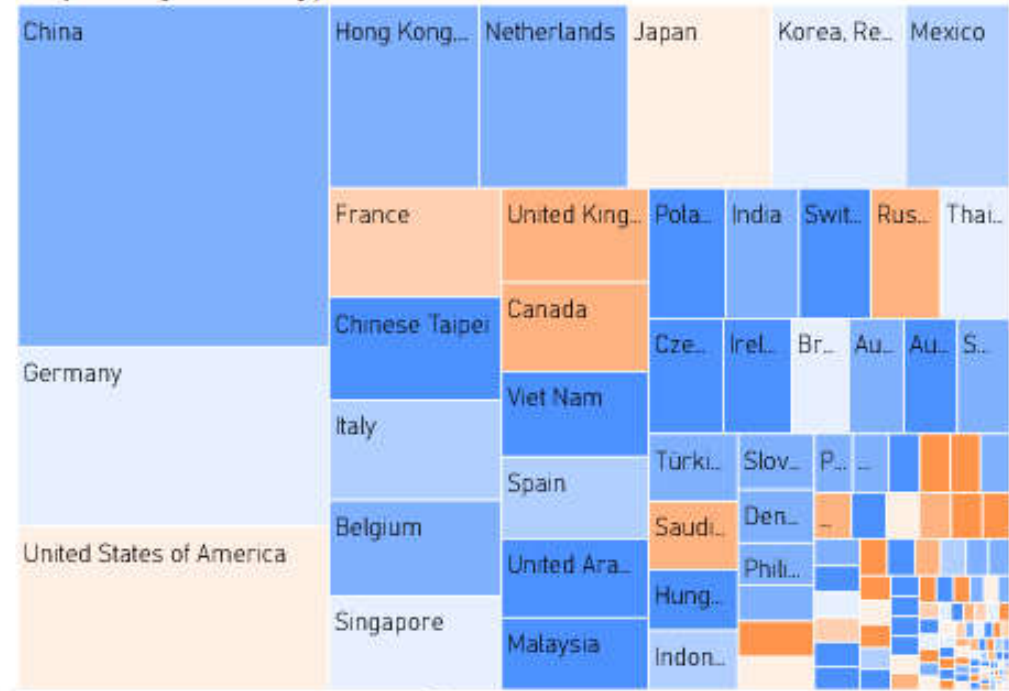
Export by country, 2020



2020 vs 2015 change:



Export by country, 2020

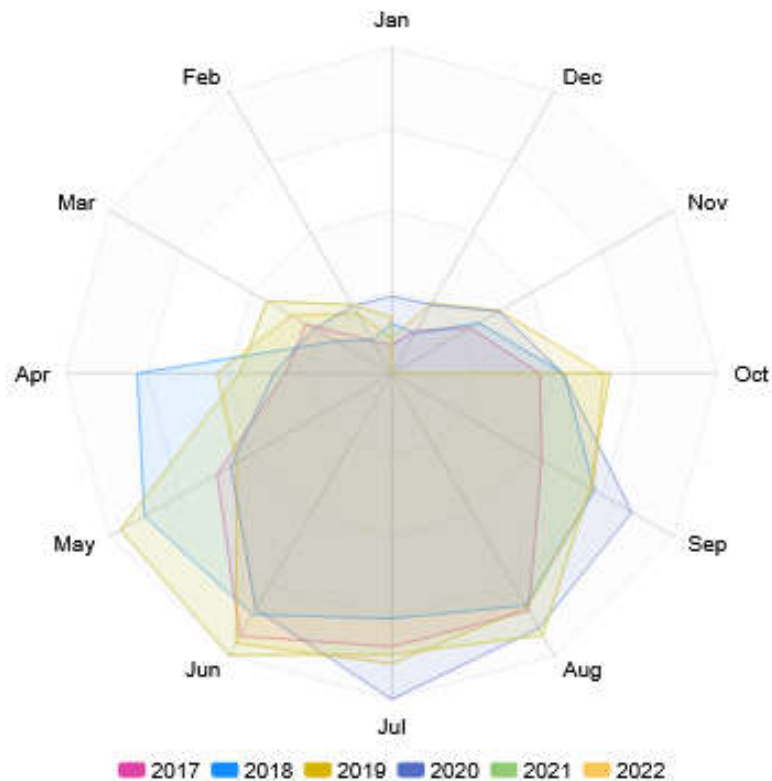


Scale: All countries

<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

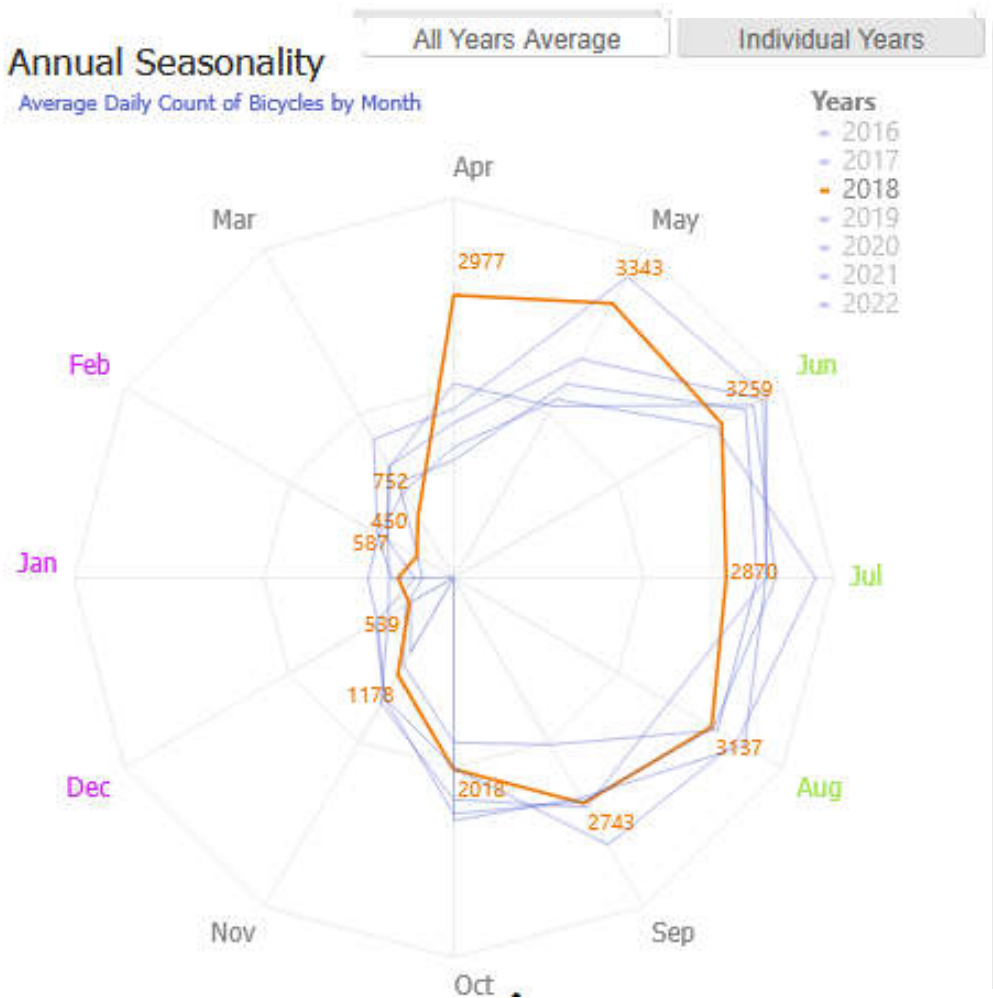
# BAD VS. GOOD CHARTS

Average Count of Bicycles by Month



Annual Seasonality

Average Daily Count of Bicycles by Month



<https://powerofbi.org/bad-and-good-charts-in-power-bi/>

---

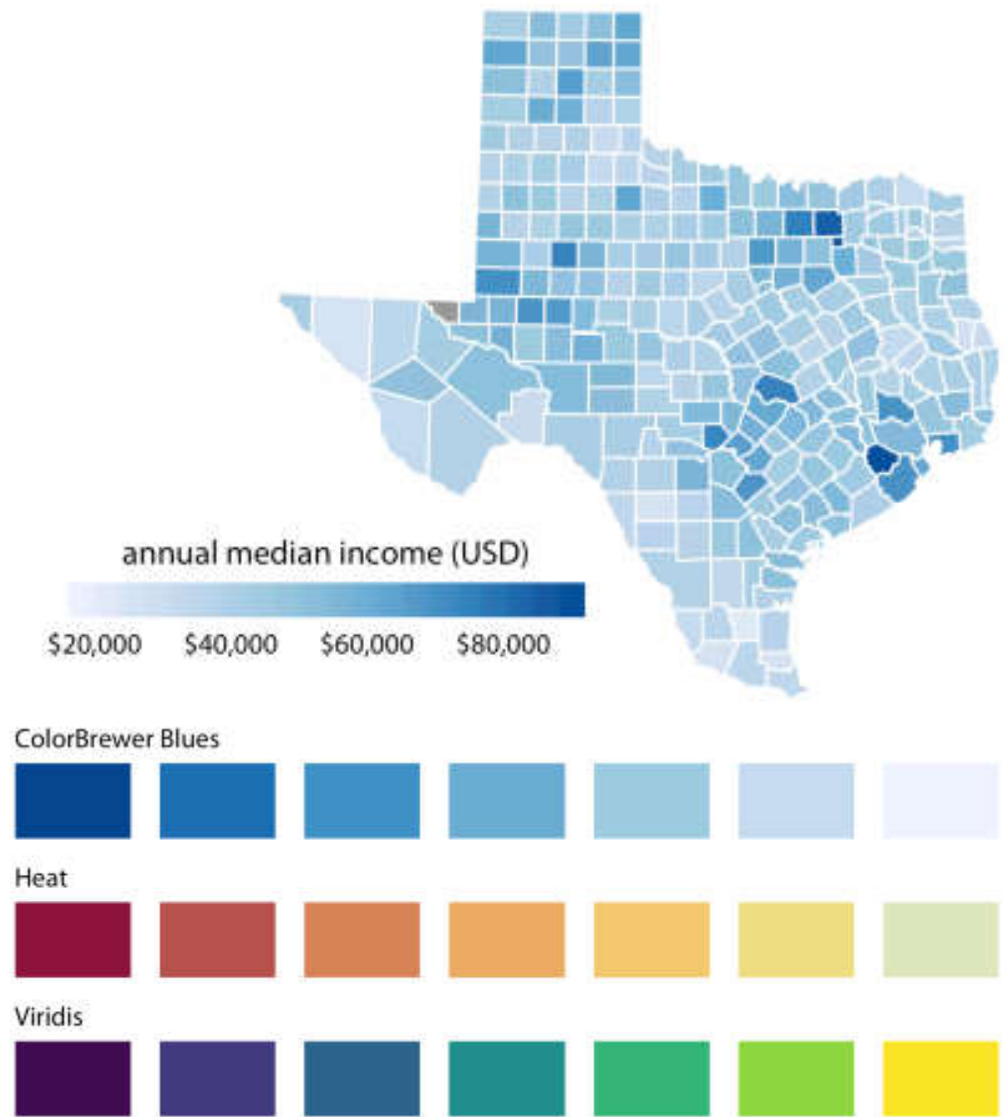
# COLOR SCALE

- To represent data values
- To distinguish groups
- To highlight



# COLOR TO REPRESENT VALUES

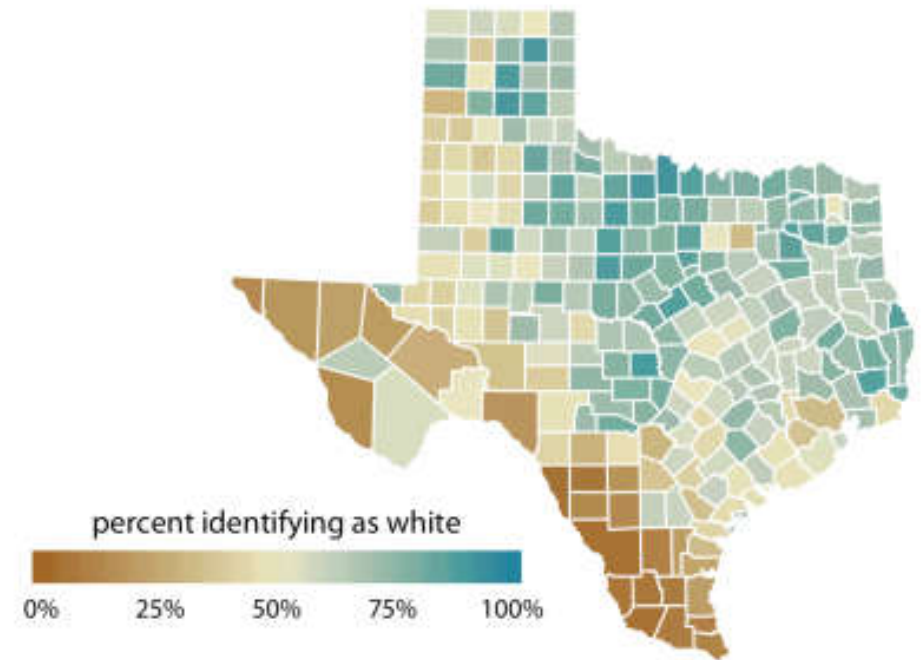
- Quantitative data values
  - E.g., income, temperature, speed
- Sequential color scale
  - Clearly indicates which values are larger or smaller than which other ones.
  - How distant two specific values are from each other.
  - The color scale needs to be perceived to vary uniformly across its entire range.





# COLOR TO REPRESENT VALUES

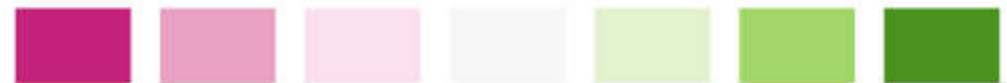
- Quantitative data values
  - E.g., income, temperature, speed
- Diverging color scale
  - One of two directions relative to a neutral midpoint.
  - How far in either direction it deviates from the midpoint.
  - The midpoint normally is a light color



CARTO Earth



ColorBrewer PiYG

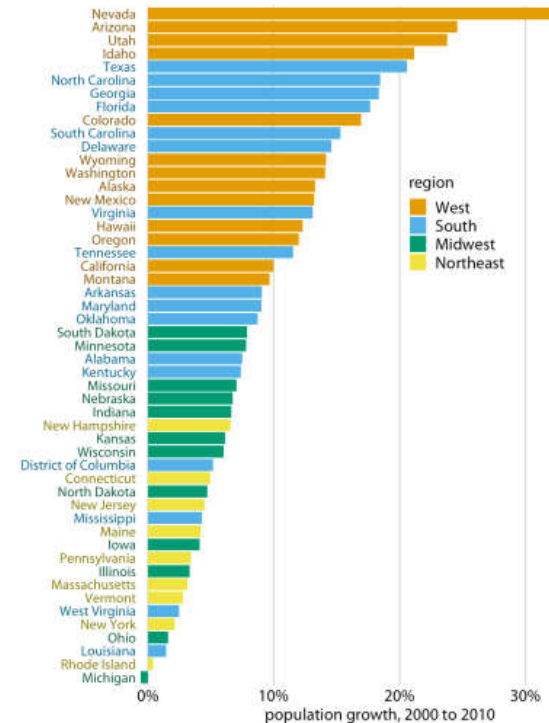


Blue-Red



# COLOR TO DISTINGUISH GROUPS

- Groups without order
  - E.g., countries on a map, manufacturers of products
- Qualitative color scale
  - Specific colors to look clearly distinct from each other while also being equivalent to each other.
  - No one color should stand out relative to the others.
  - The colors should not create the impression of an order (e.g., successively lighter)



Okabe Ito



ColorBrewer Dark2

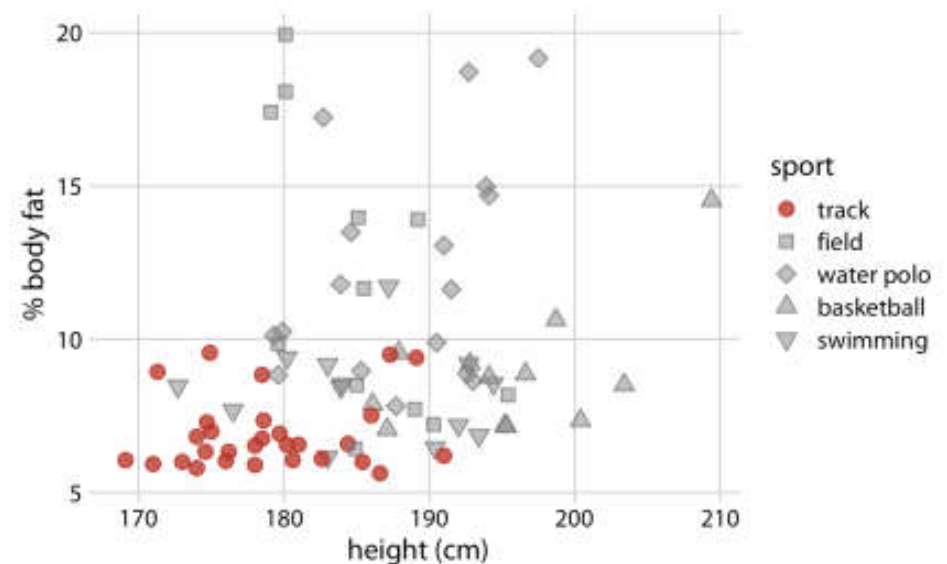


ggplot2 hue



# COLOR TO HIGHLIGHT

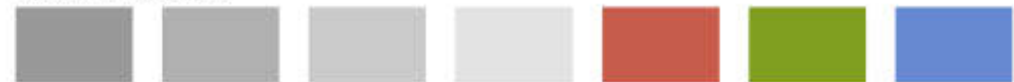
- To highlight specific elements in the data.
- Accent color scale
  - Vividly stands out against the rest of the figure.
  - Contains both a set of subdued colors and a matching set of stronger, darker, and/or more saturated colors.
- It is critical that the baseline colors do not compete for attention.



Okabe Ito Accent



Grays with accents

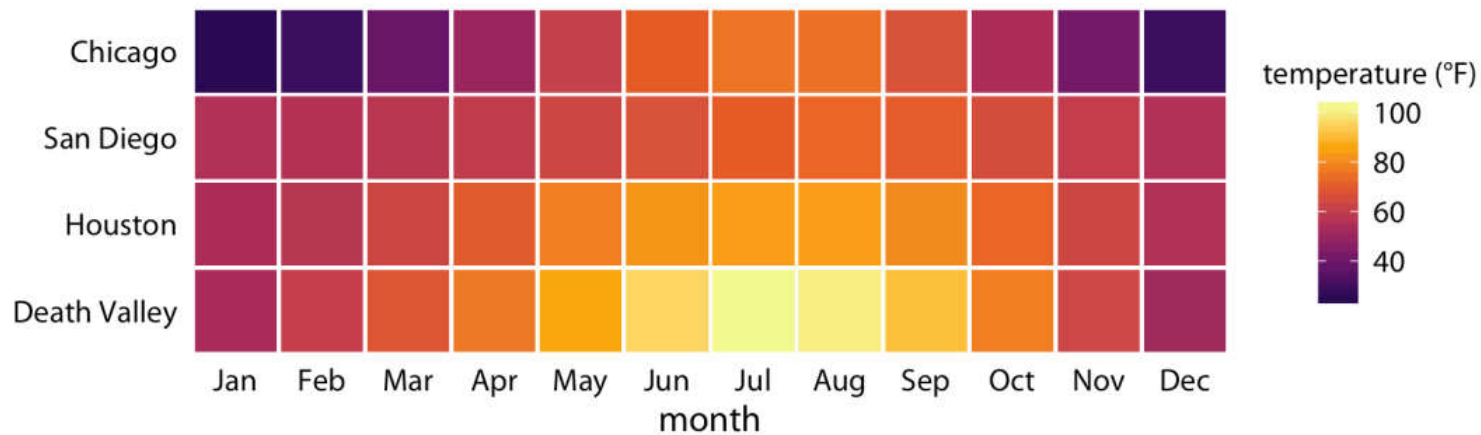
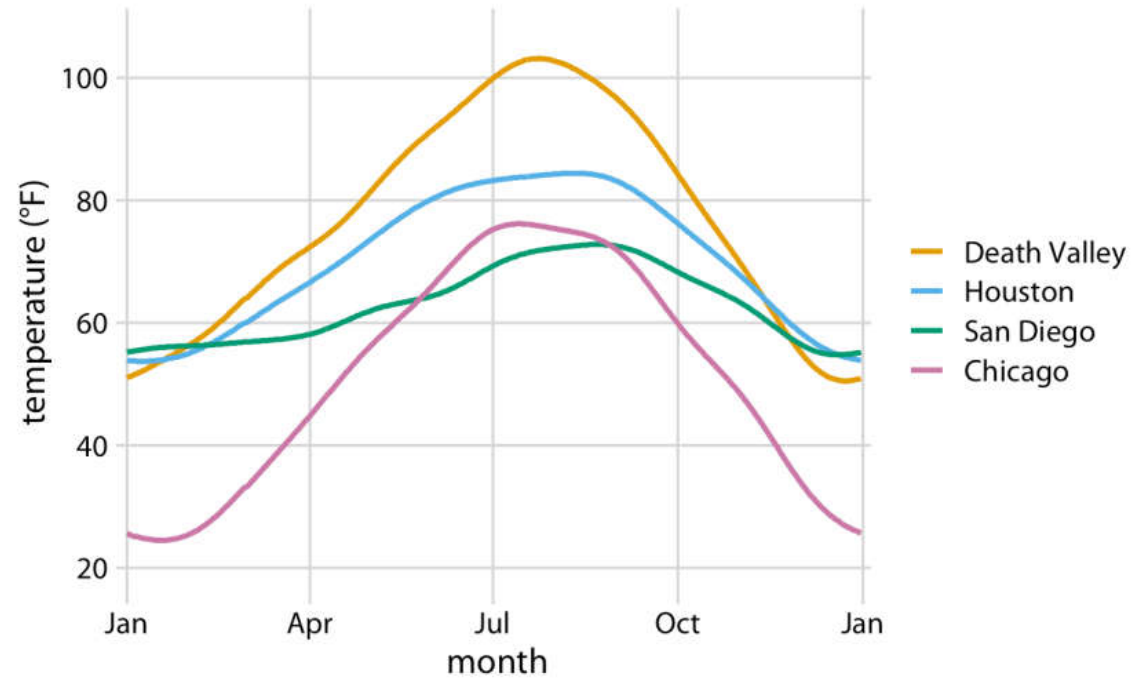


ColorBrewer Accent



# FOR INSTANCE

Month	Day	Location	Station ID	Temperature (°F)
Jan	1	Chicago	USW00014819	25.6
Jan	1	San Diego	USW00093107	55.2
Jan	1	Houston	USW00012918	53.9
Jan	1	Death Valley	USC00042319	51.0
Jan	2	Chicago	USW00014819	25.5
Jan	2	San Diego	USW00093107	55.3
Jan	2	Houston	USW00012918	53.8
Jan	2	Death Valley	USC00042319	51.2



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

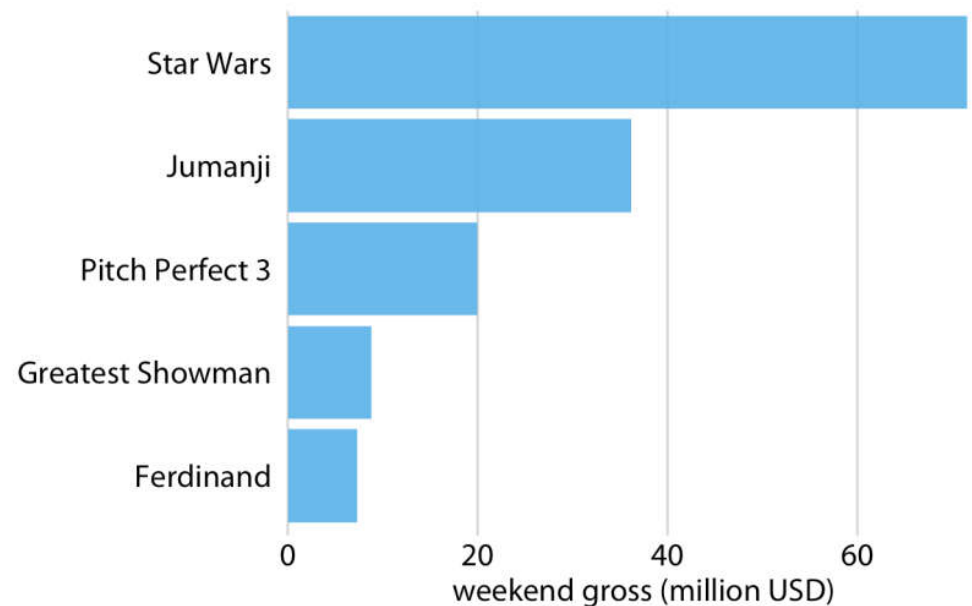
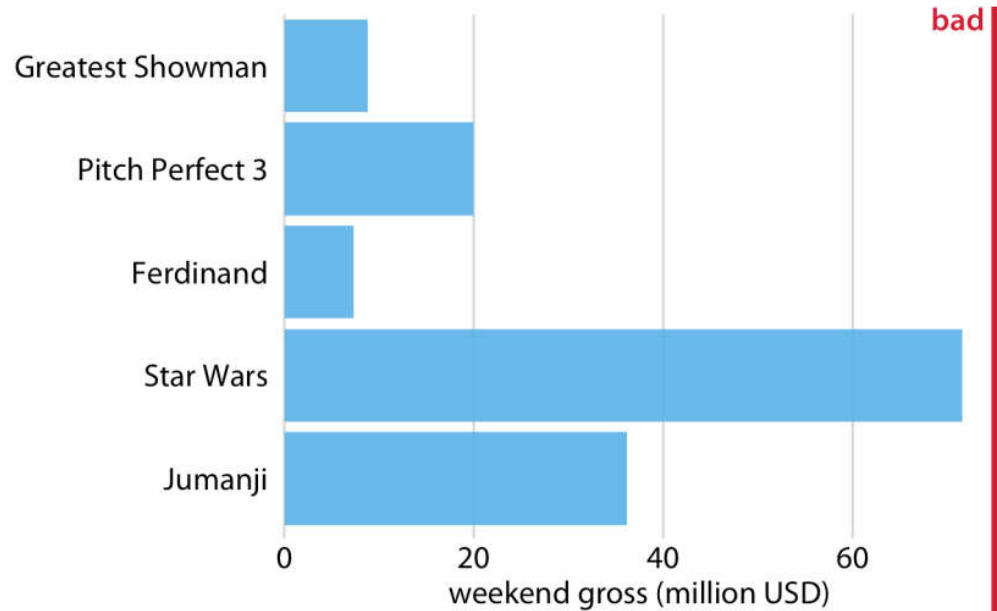
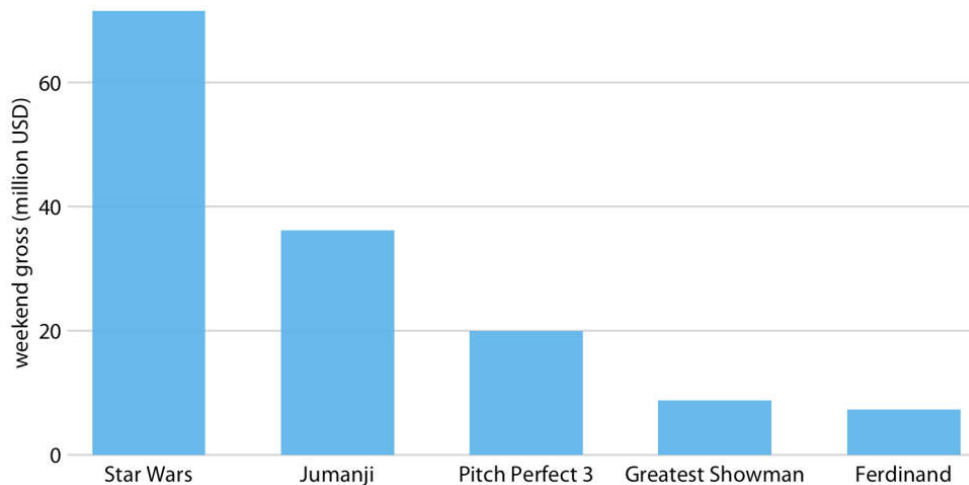
---

# VISUALIZING AMOUNTS

- When we are interested in the magnitude of some set of numbers.
  - E.g., total sales of different brands, total customer in different areas
- Standard visualization
  - Simple bars
  - Grouped bars
  - Stacked bars
- Alternatives
  - Dot plot
  - Heatmap

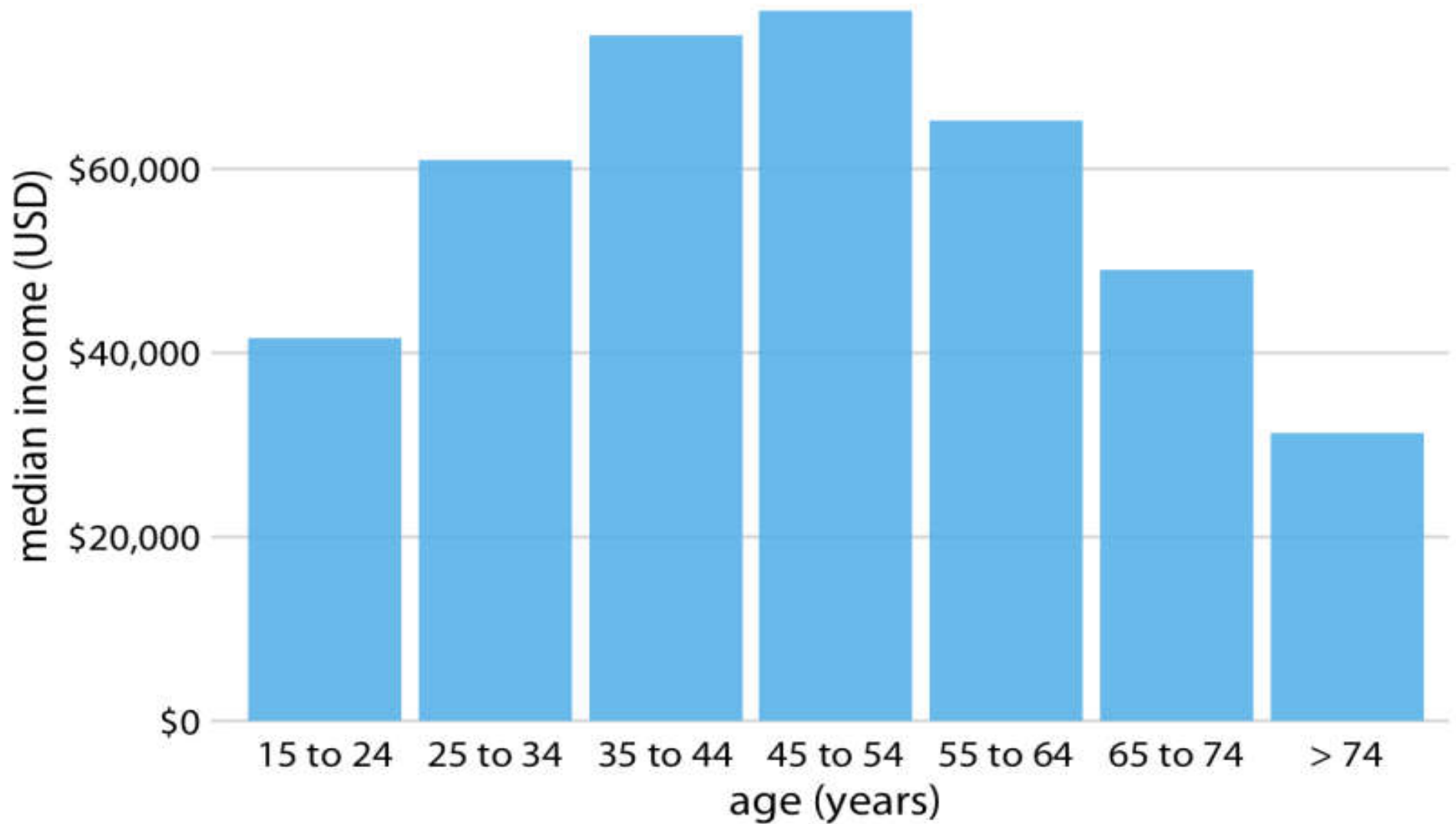
# SIMPLE BARS

Rank	Title	Weekend gross
1	<i>Star Wars: The Last Jedi</i>	\$71,565,498
2	<i>Jumanji: Welcome to the Jungle</i>	\$36,169,328
3	<i>Pitch Perfect 3</i>	\$19,928,525
4	<i>The Greatest Showman</i>	\$8,805,843
5	<i>Ferdinand</i>	\$7,316,746



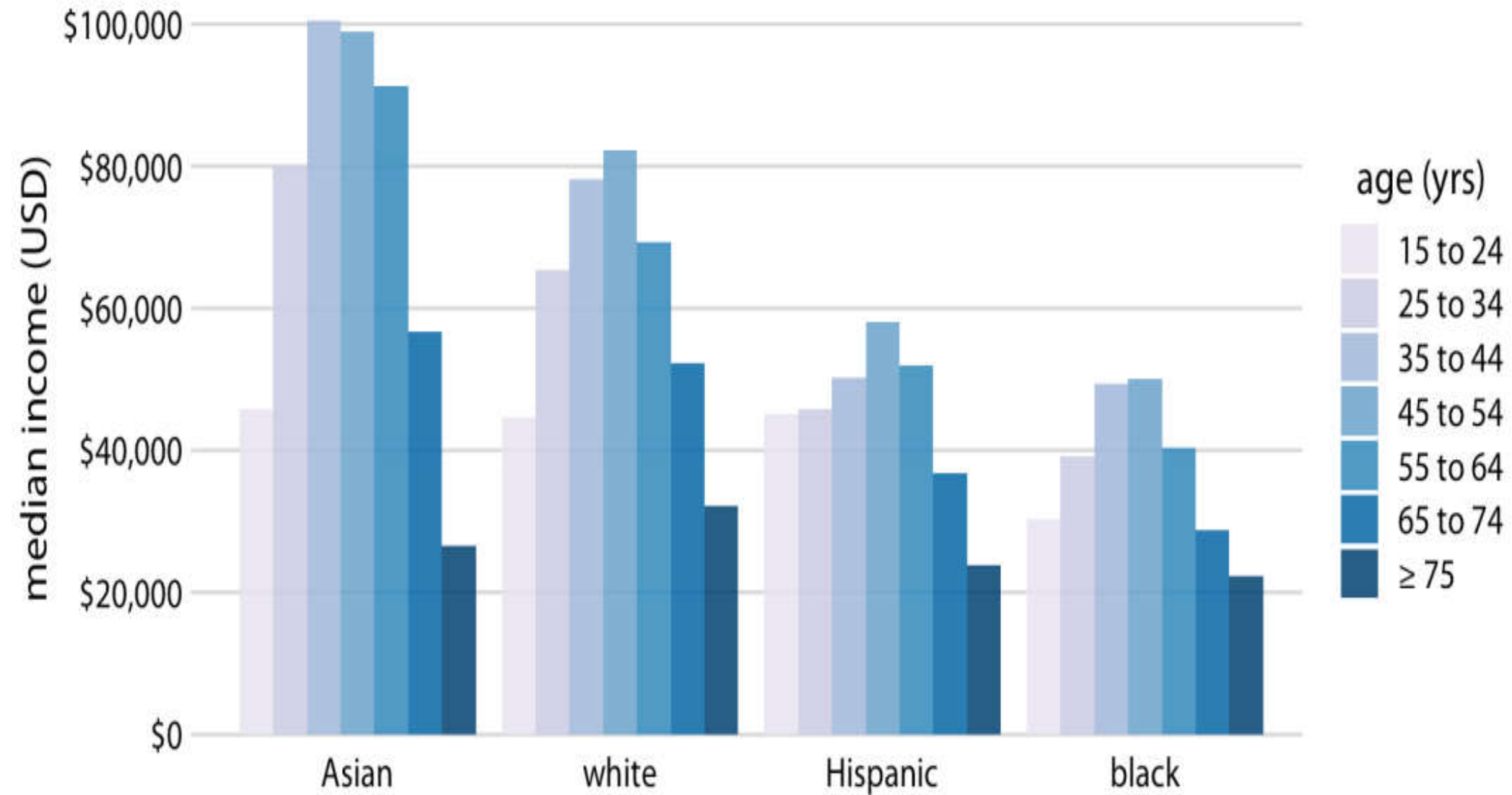
Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# SIMPLE BARS



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

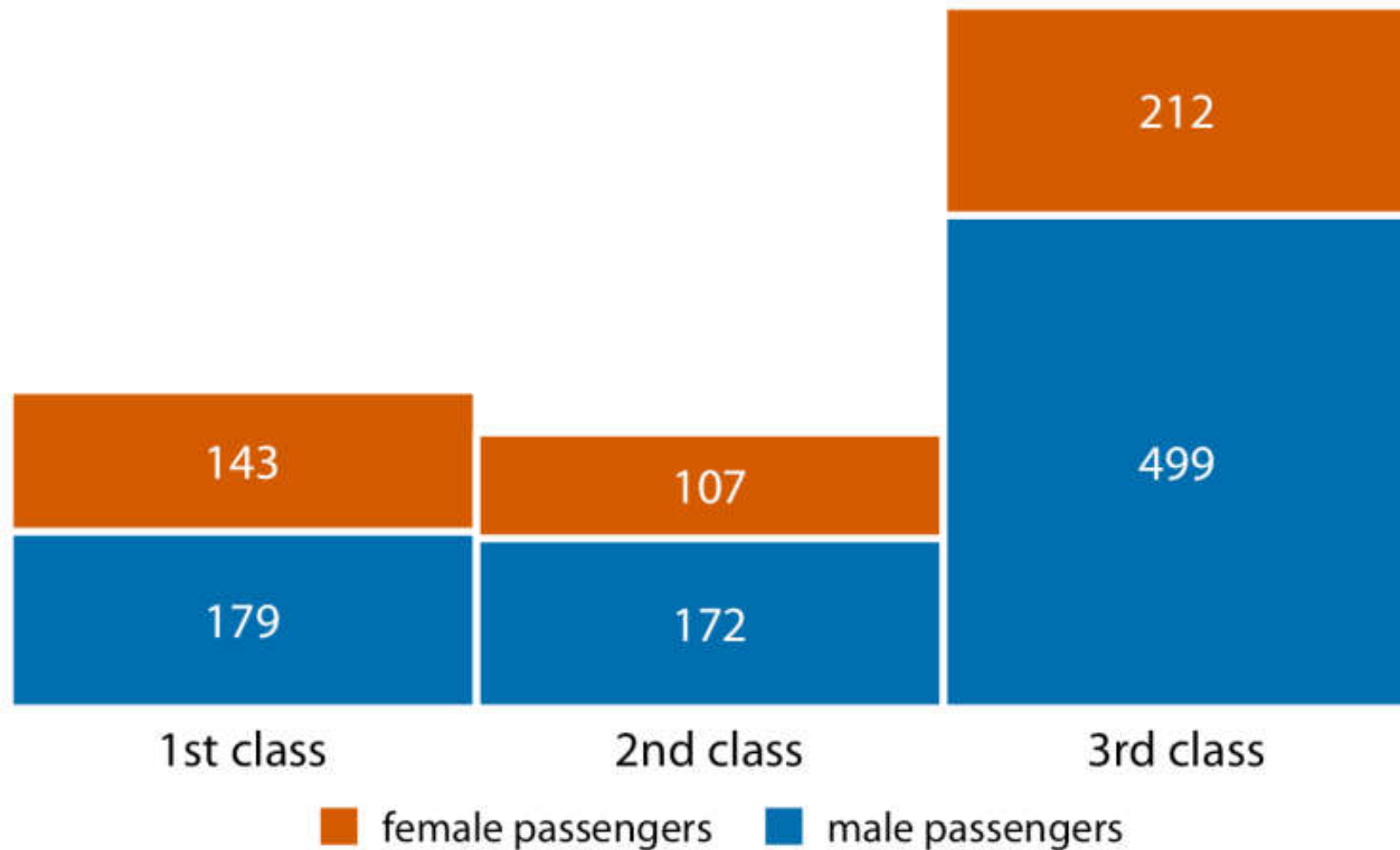
# GROUP BARS



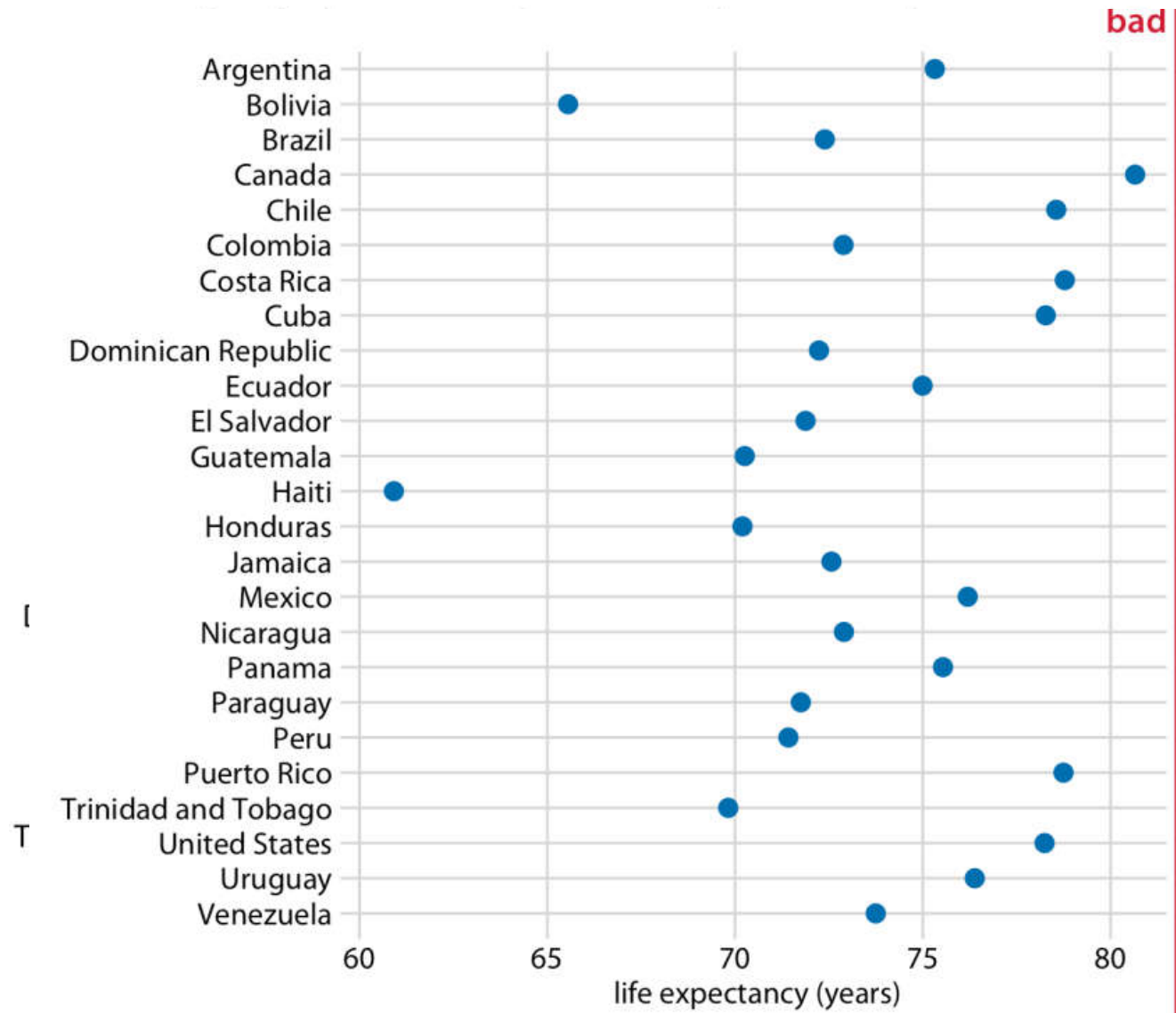
Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019



# STACKED BARS

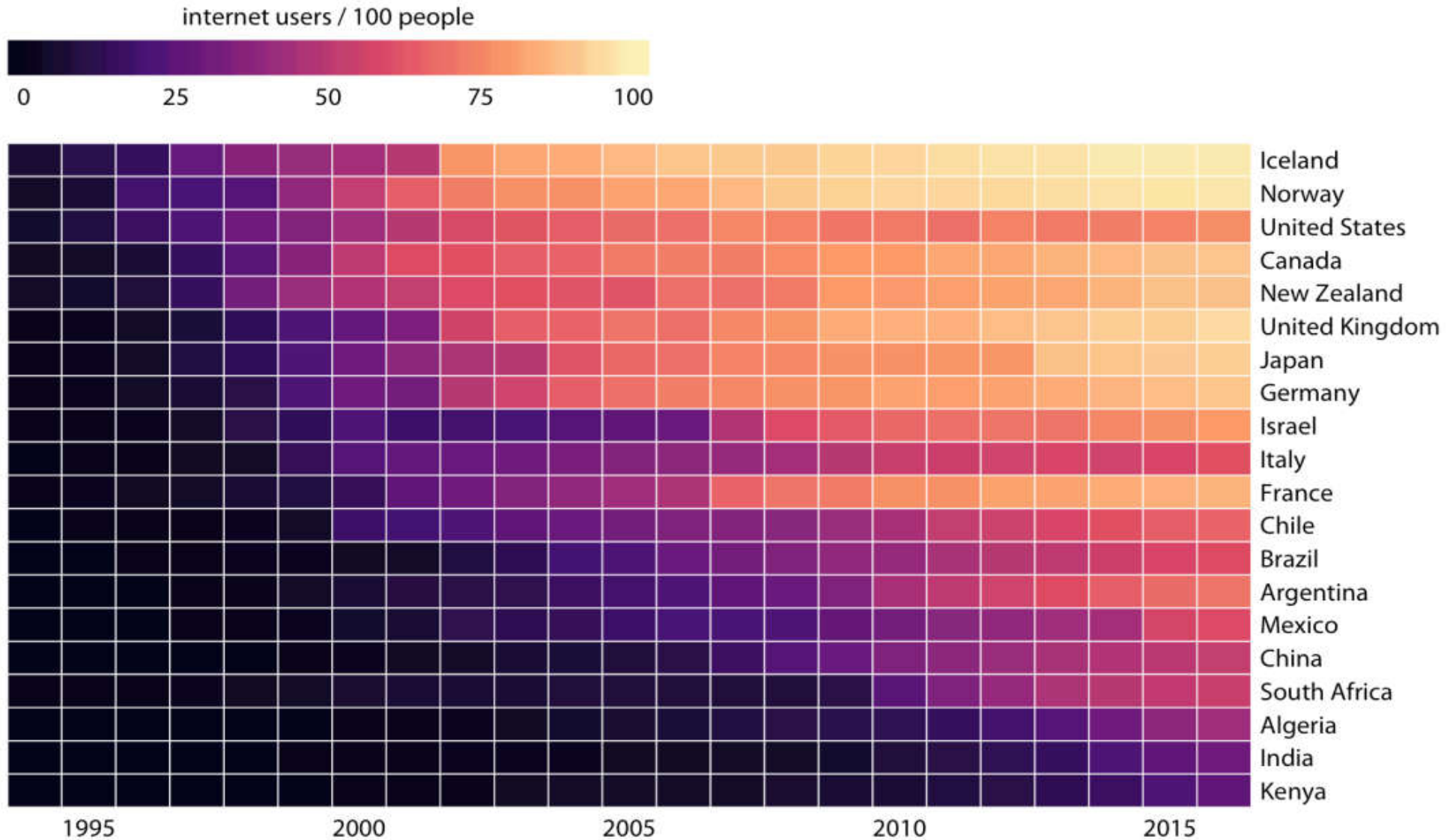


# DOT PLOTS



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# DOT PLOTS



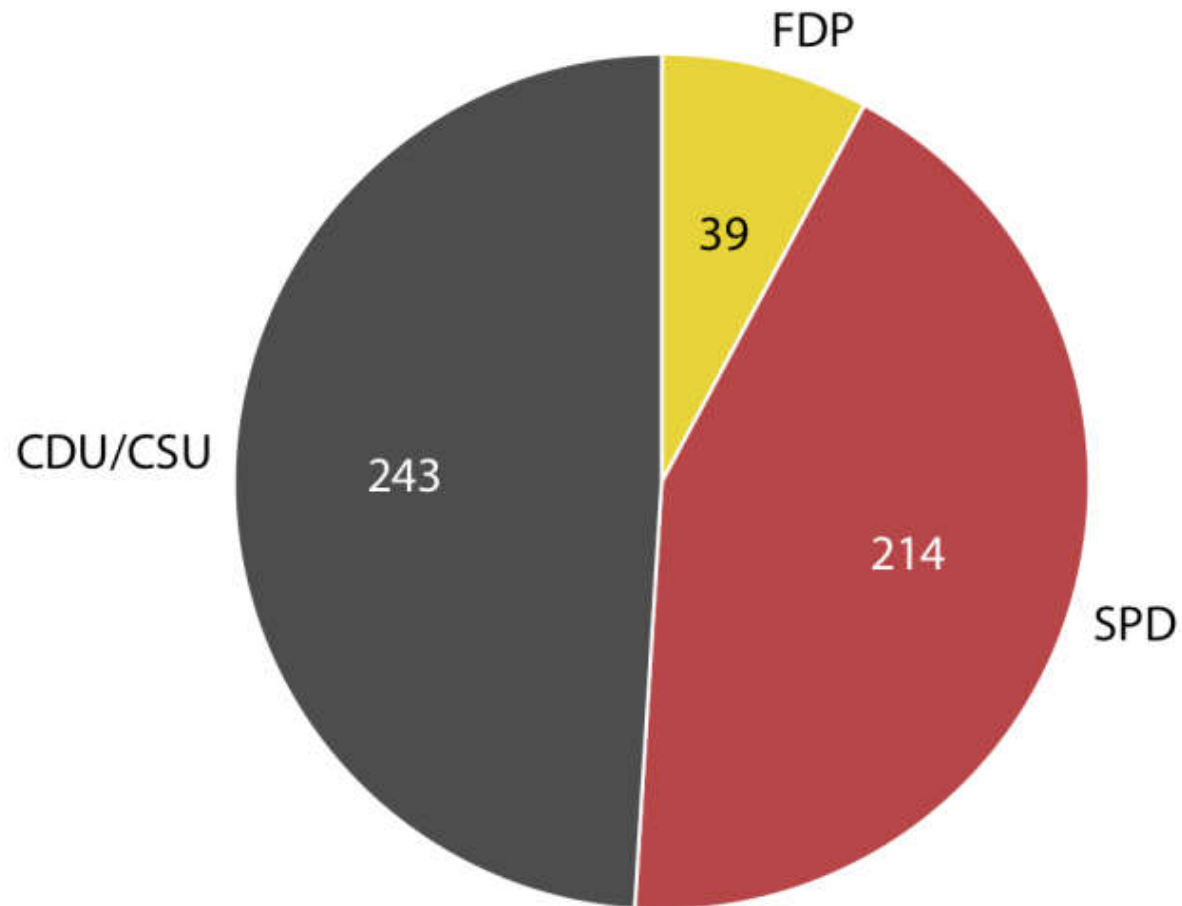
Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

---

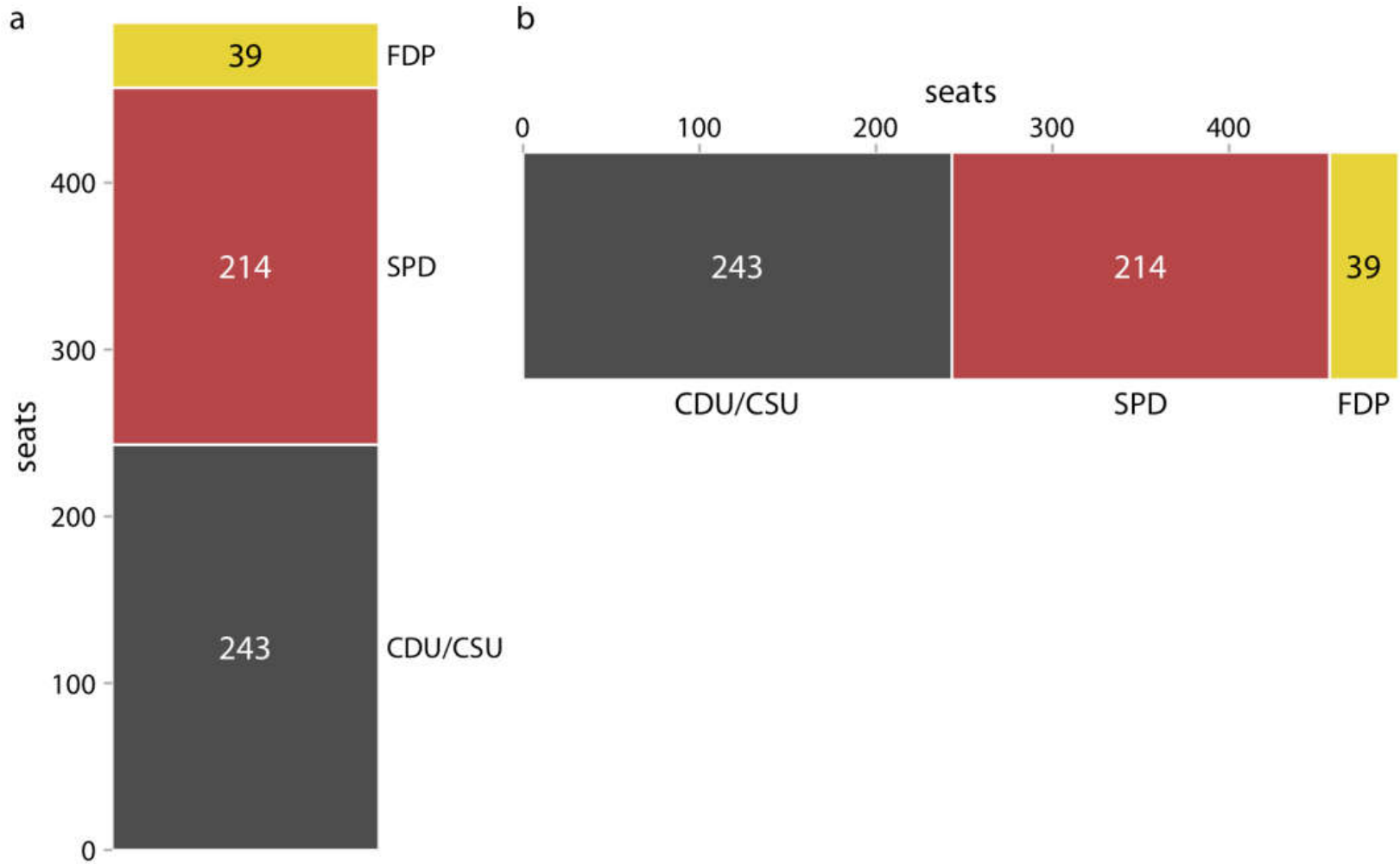
# VISUALIZING PROPORTIONS

- How some group, entity, or amount breaks down into individual pieces that each represent a proportion of the whole.
  - ❑ Pie charts
  - ❑ Stacked bars
  - ❑ Stacked densities

# PIE CHARTS

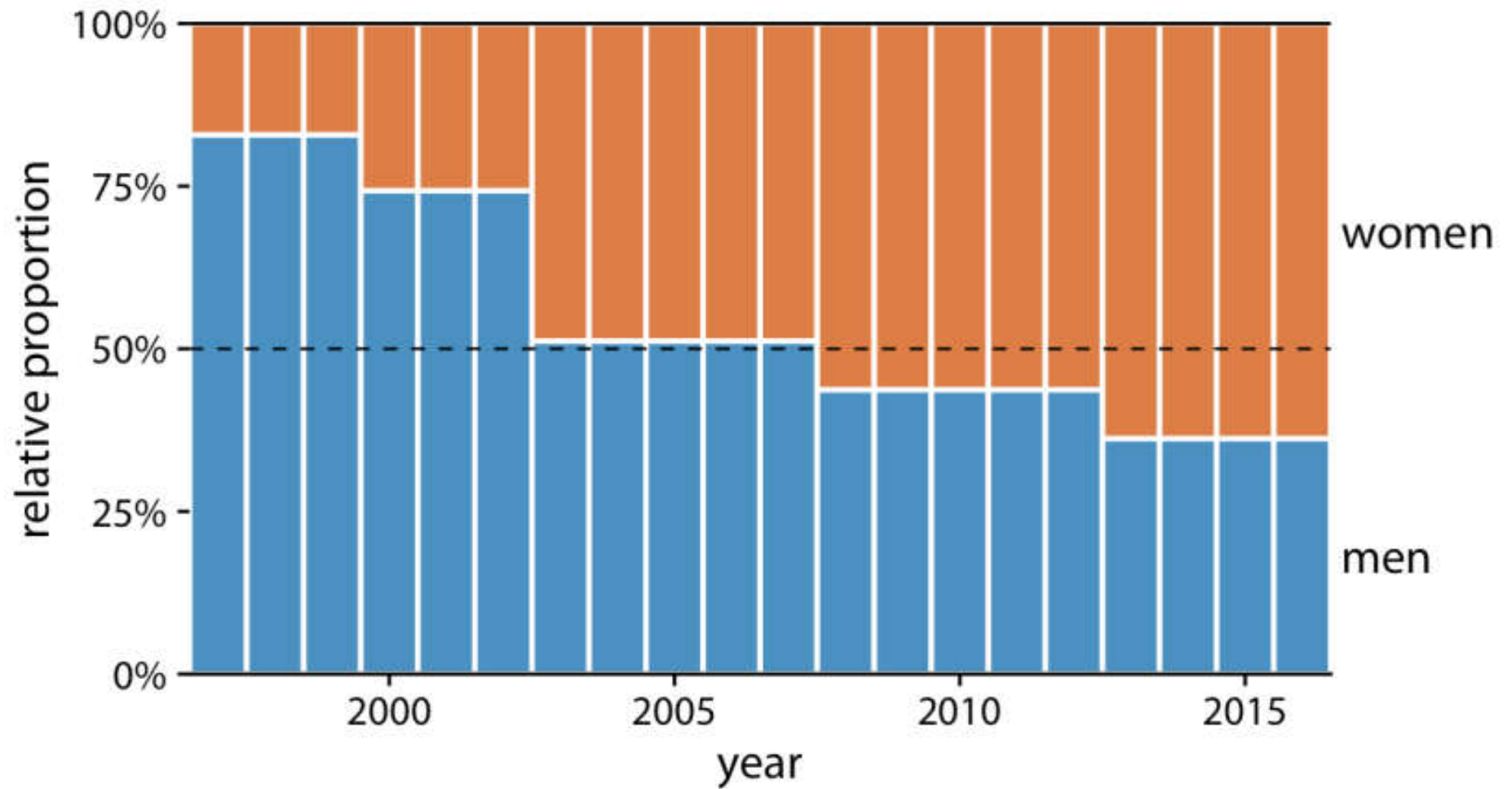


# STACKED BARS

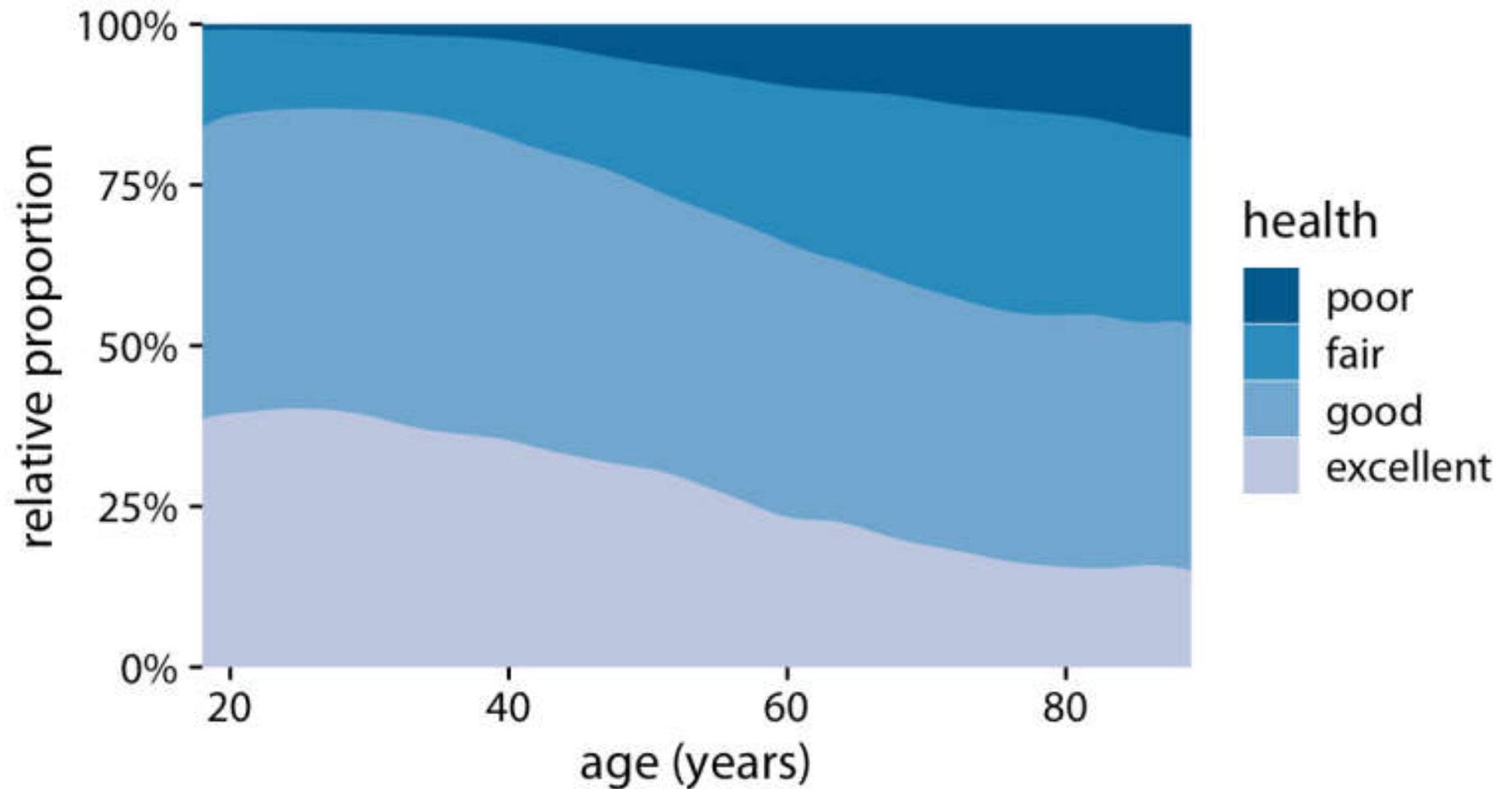


Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# STACKED BARS

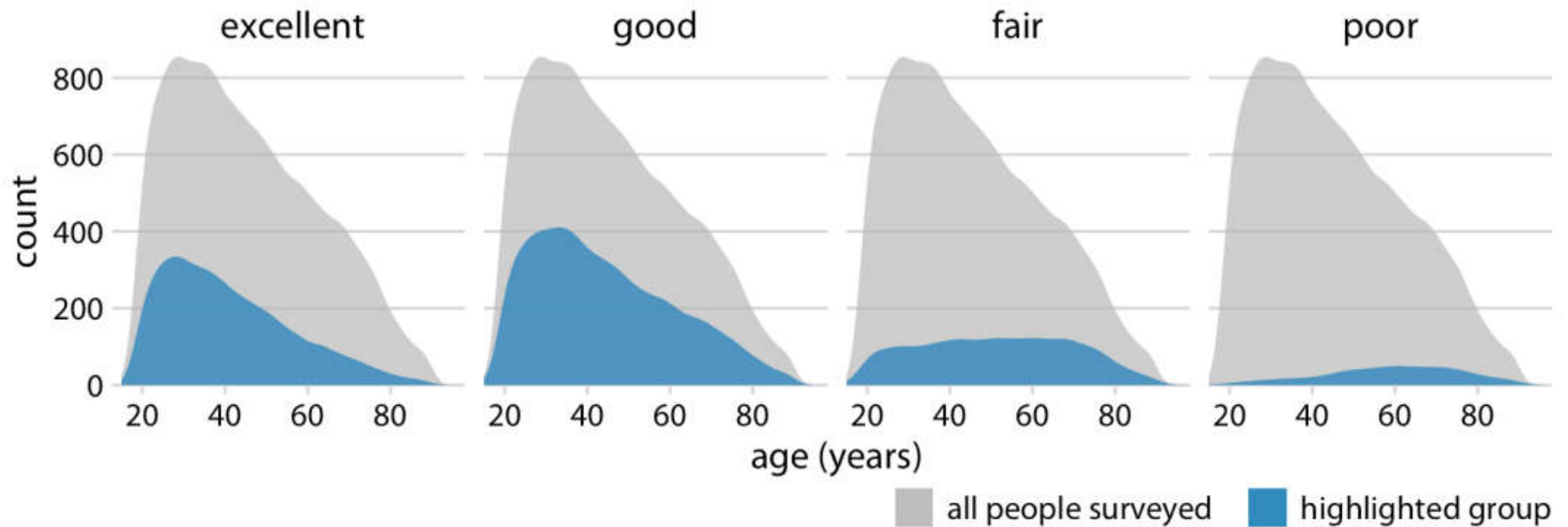


# STACKED DENSITIES





# STACKED DENSITIES

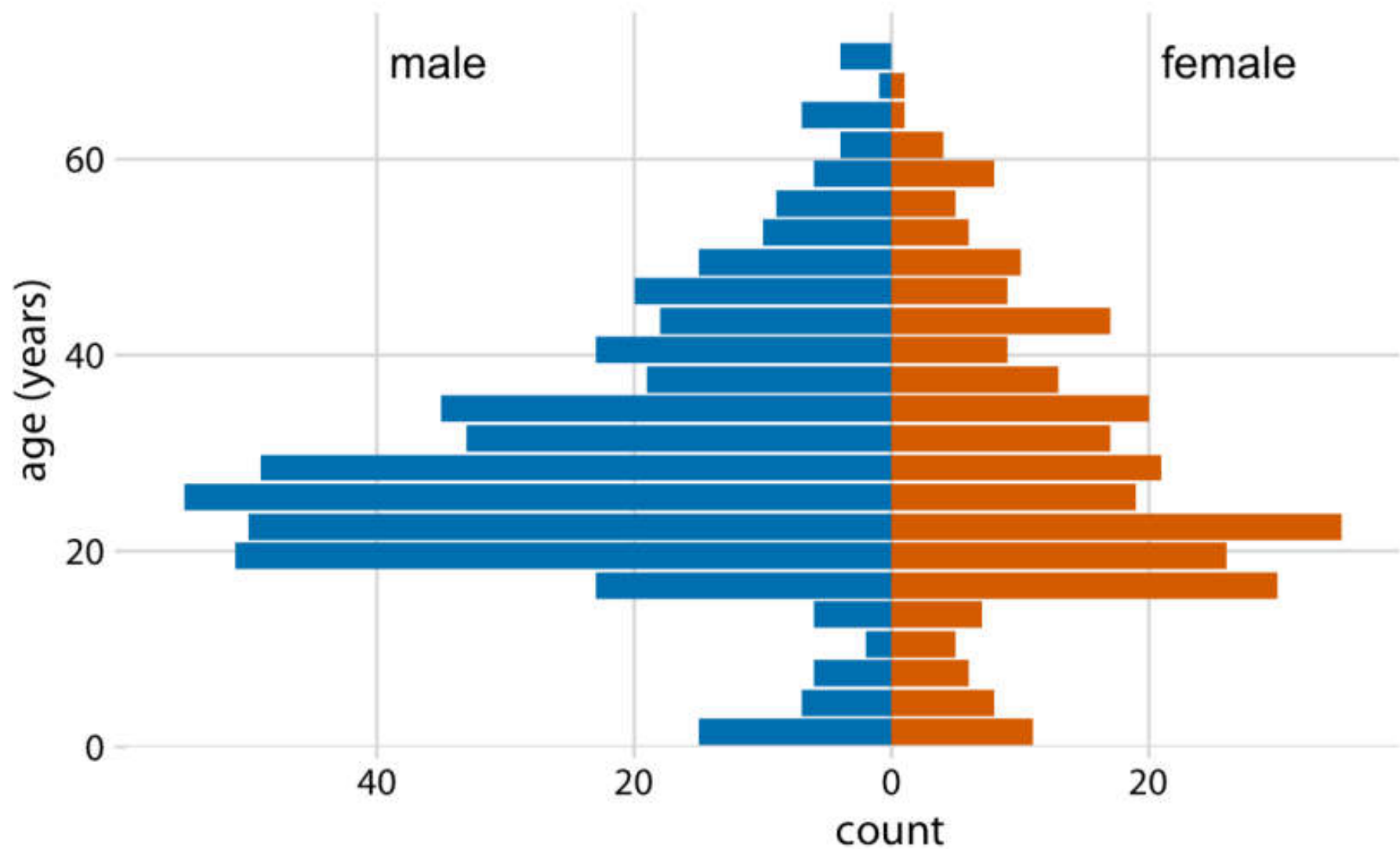


---

# VISUALIZING DISTRIBUTIONS

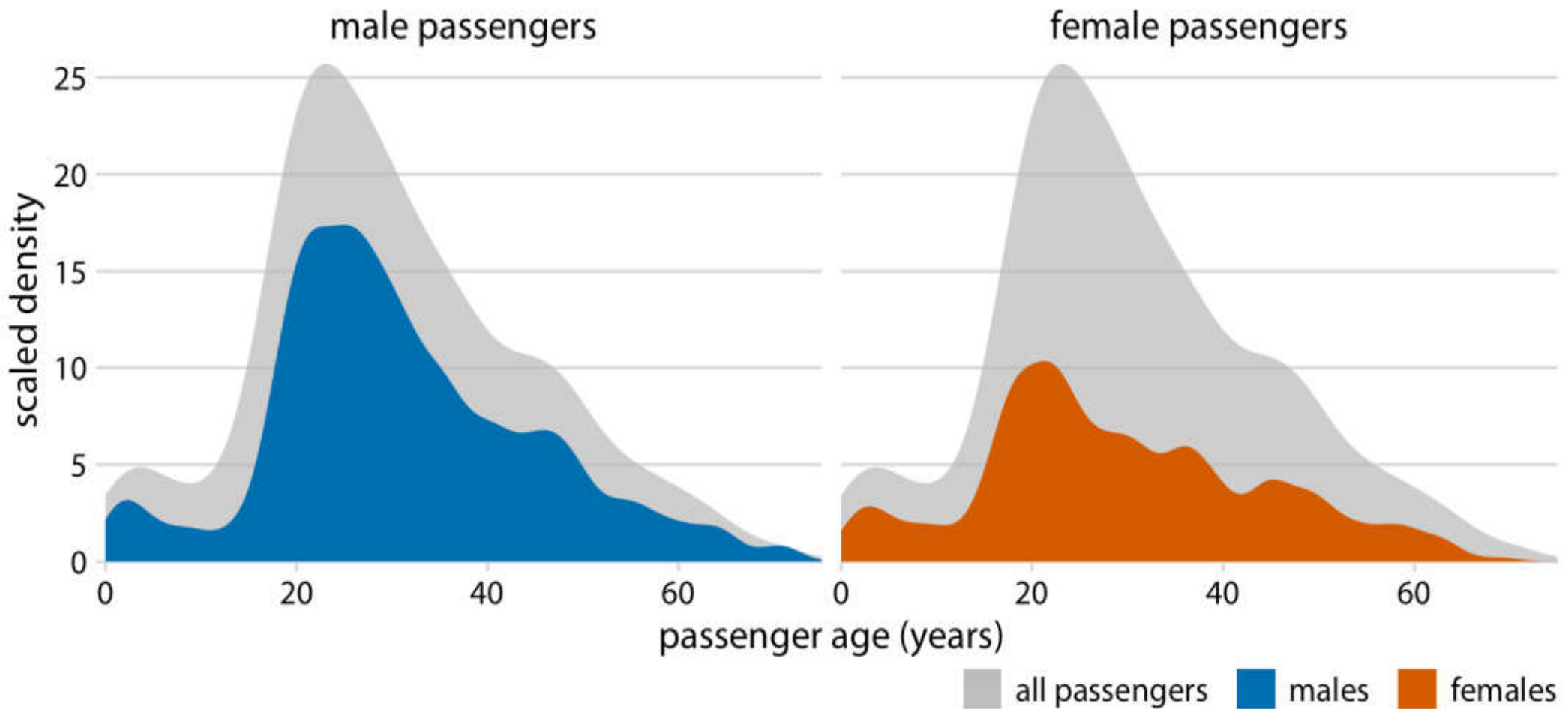
- To understand how a particular variable is distributed in a dataset.
  - Histograms
  - Density plots

# HISTOGRAMS



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# DENSITY PLOTS



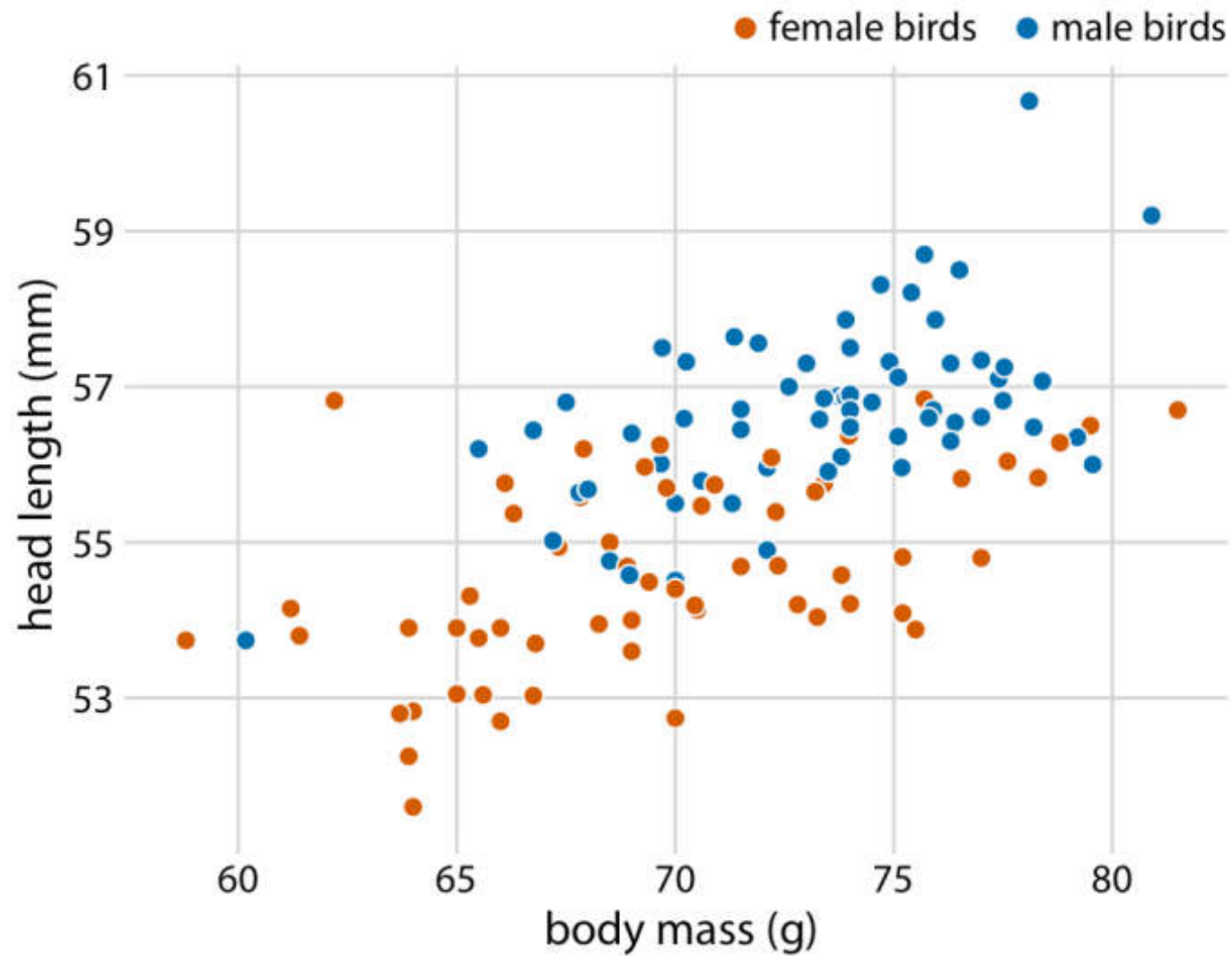
Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

---

# VISUALIZING ASSOCIATIONS

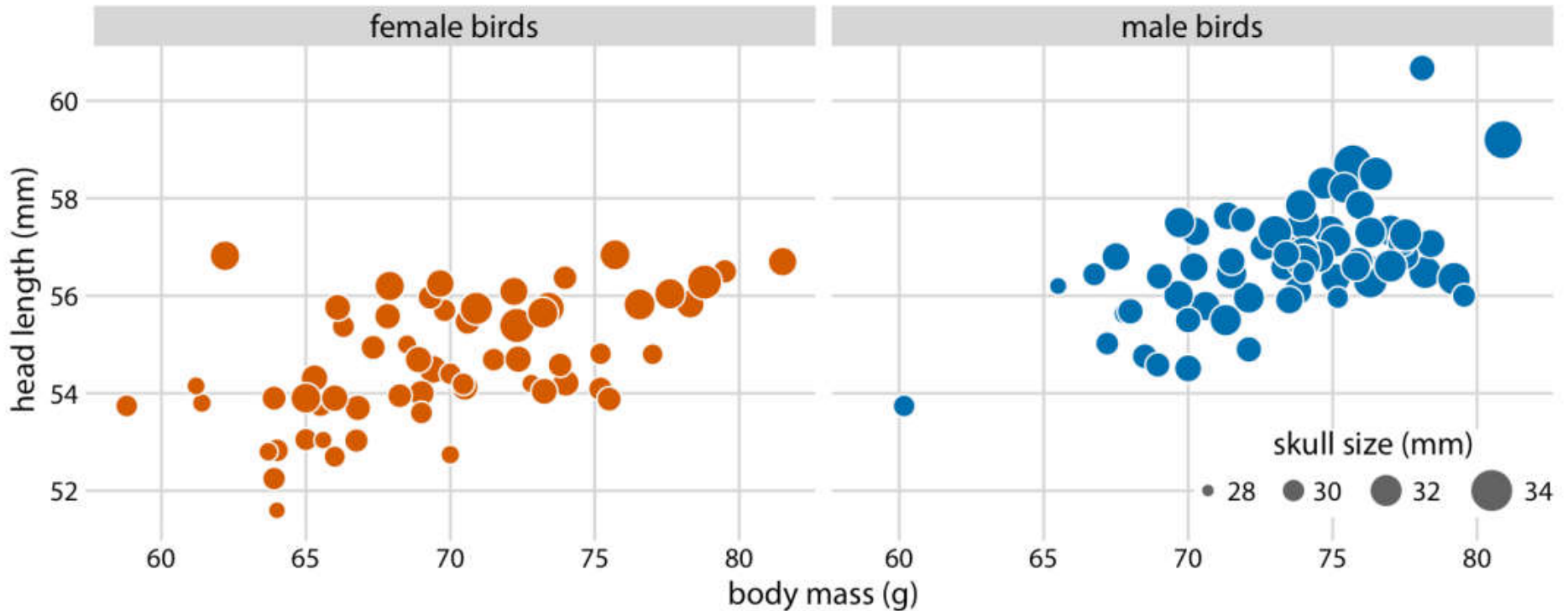
- How these variables relate to each other
  - E.g., height and weight
- Report
  - Scatterplot
  - Bubble chart
  - Scatterplot matrix
  - Correlogram

# SCATTERPLOT

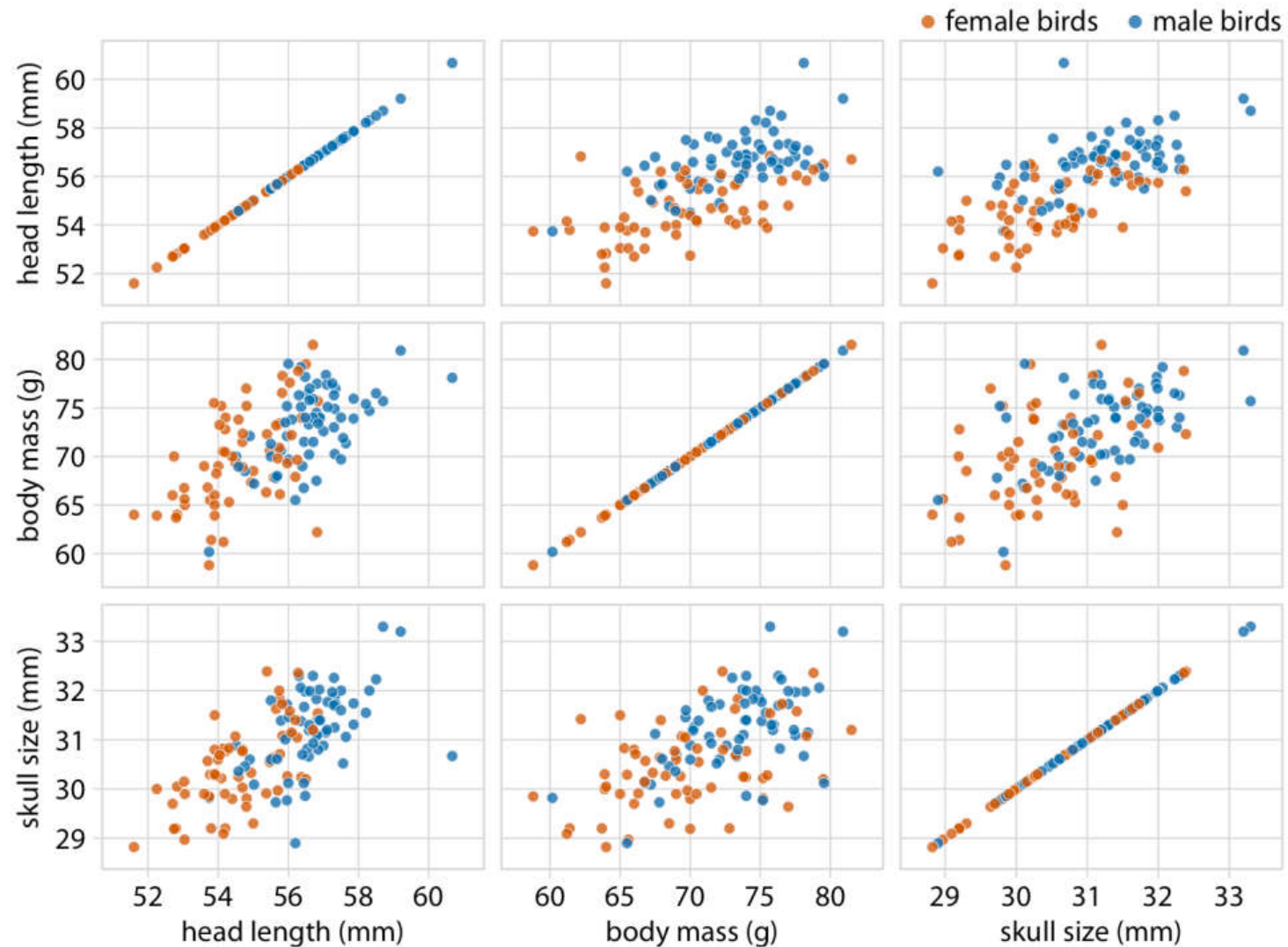


Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# BUBBLE CHART



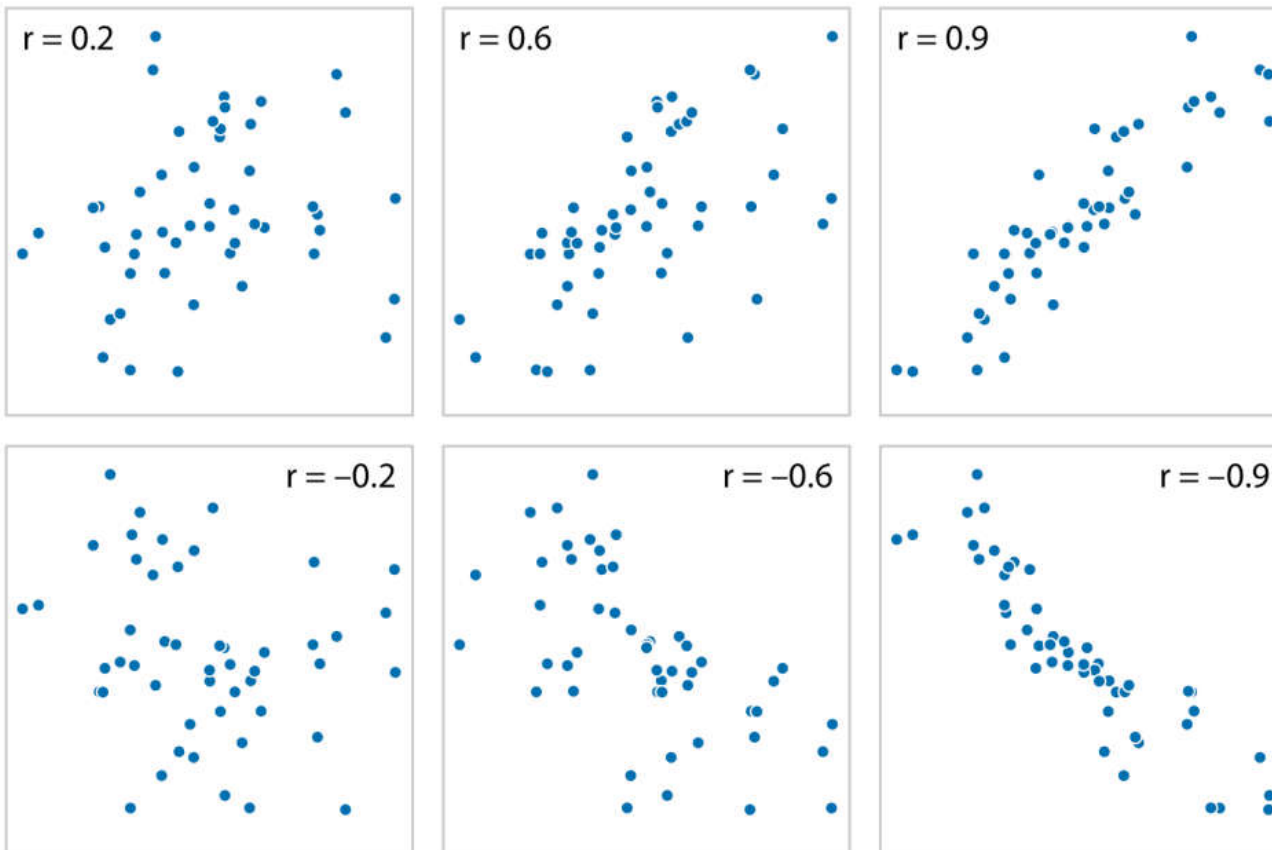
# SCATTERPLOT MATRIX





# CORRELOGRAMS

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \sqrt{\sum_i (y_i - \bar{y})^2}}$$

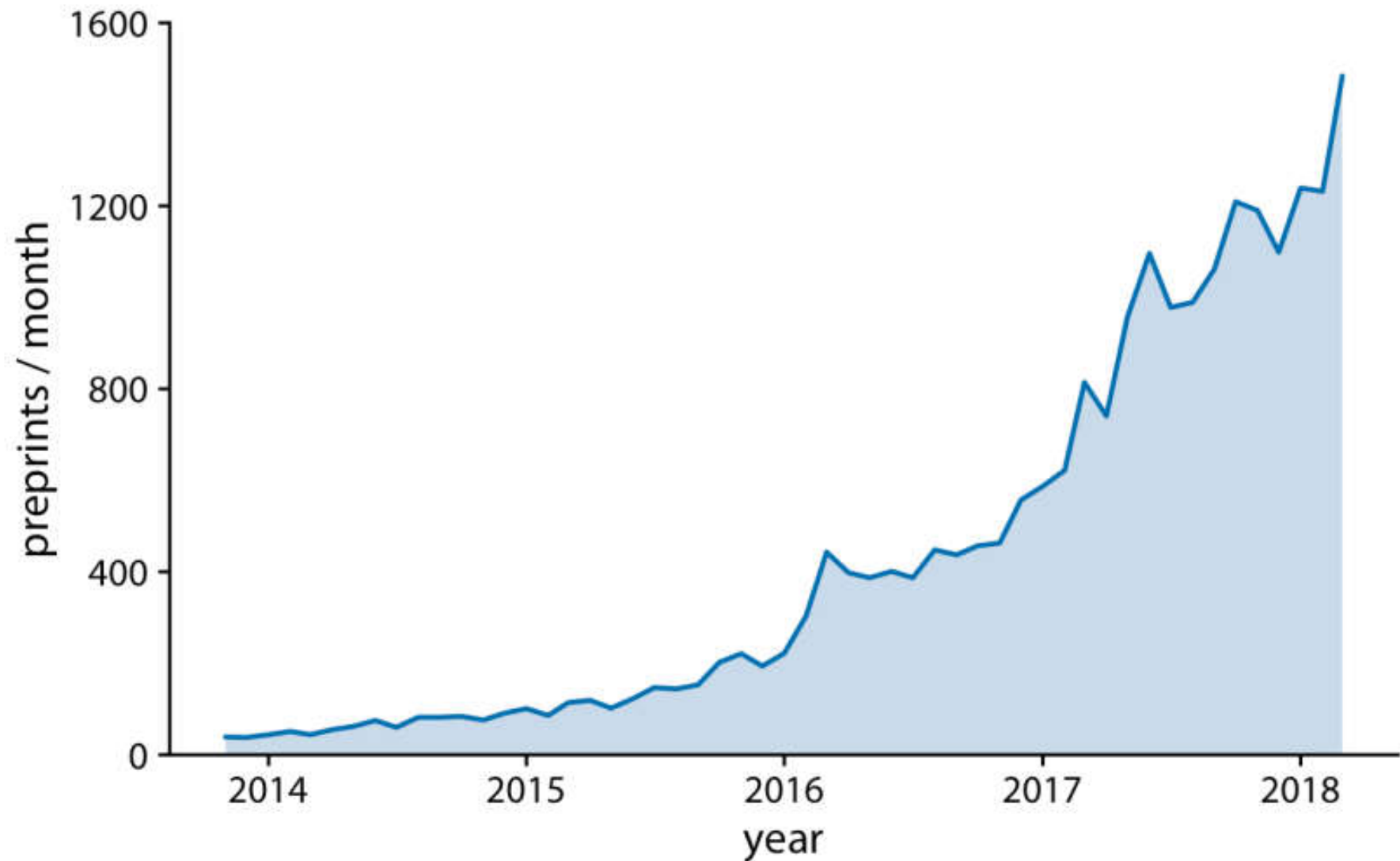


---

# VISUALIZING TIME SERIES

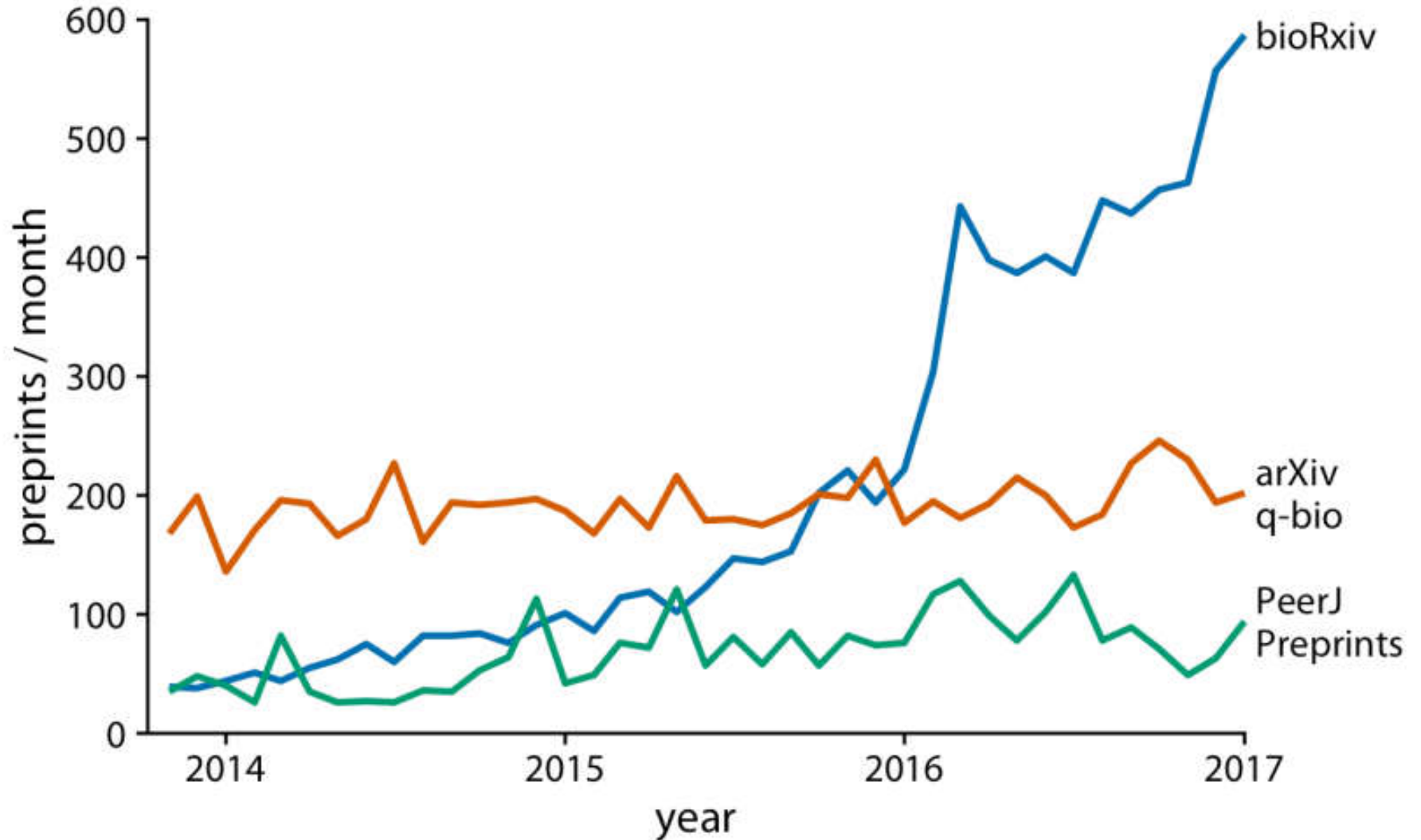
- To see pattern by time
  - Single time series
  - Multiple time series
  - Time series with two or more variables

# SINGLE TIME SERIES



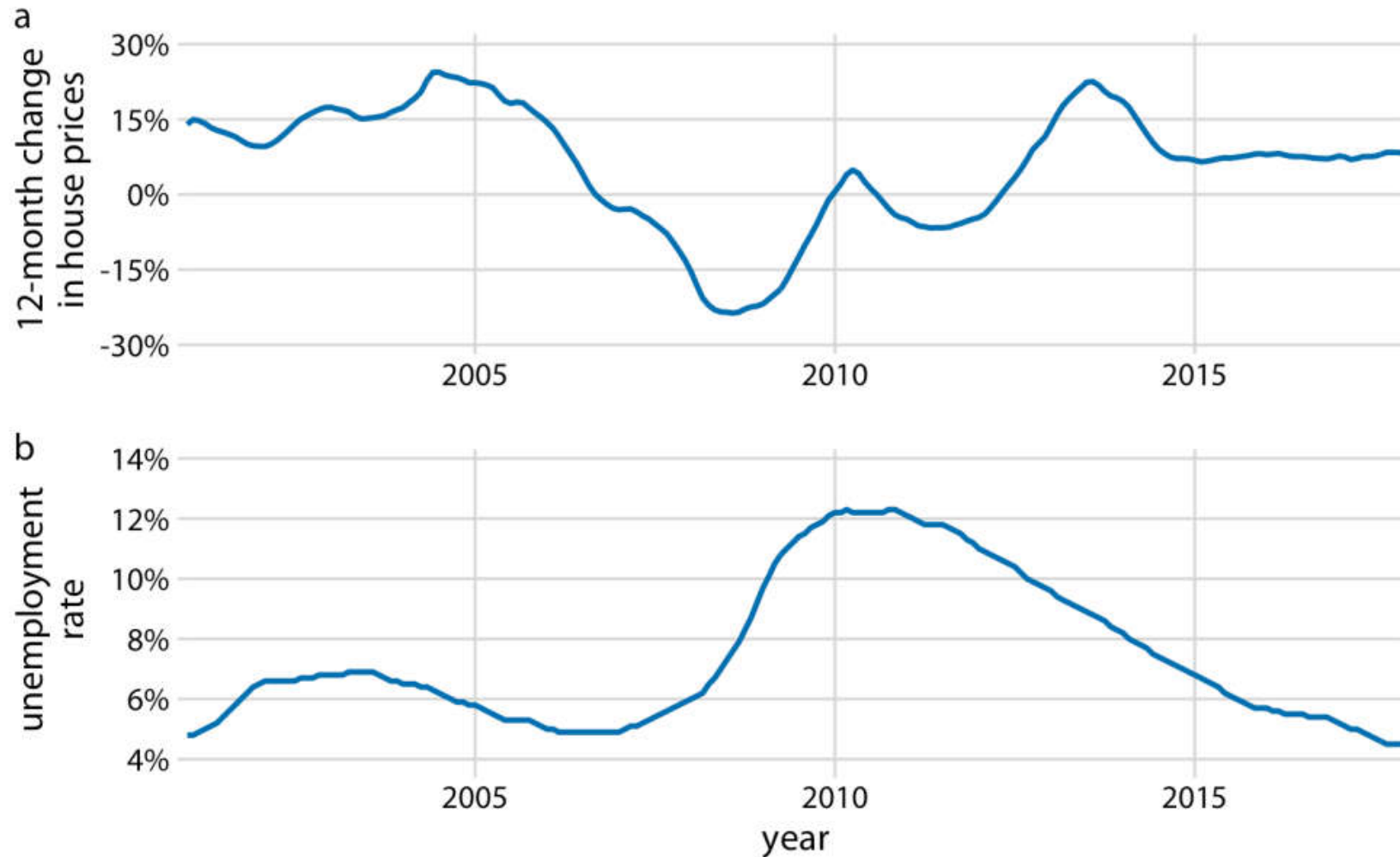
Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# MULTIPLE TIME SERIES



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# TIME SERIES WITH TWO VARIABLES

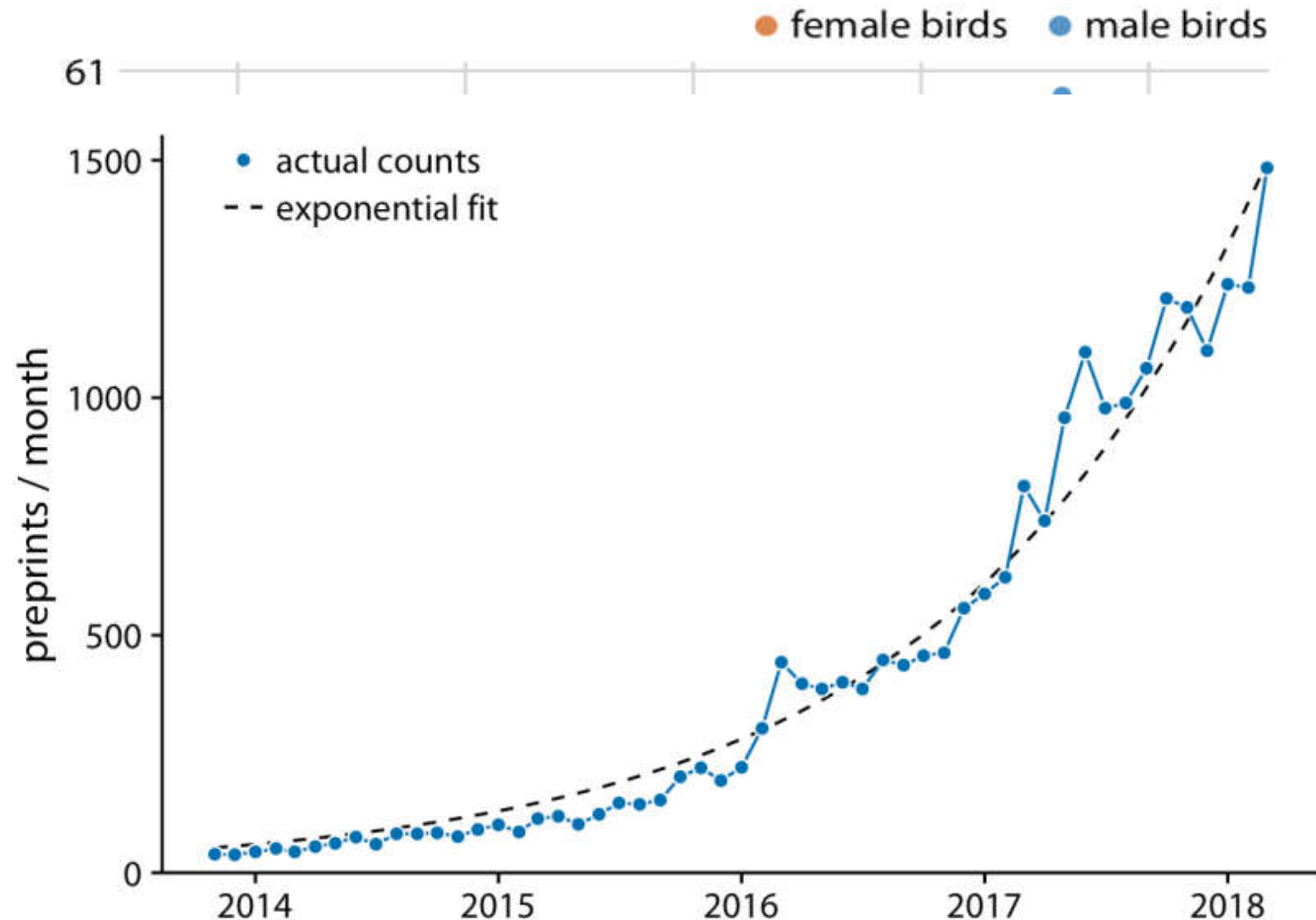


---

# VISUALIZING TRENDS

- To see key features of data
  - Smoothing
  - Linear
  - Curve

# TIME SERIES WITH TWO VARIABLES



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

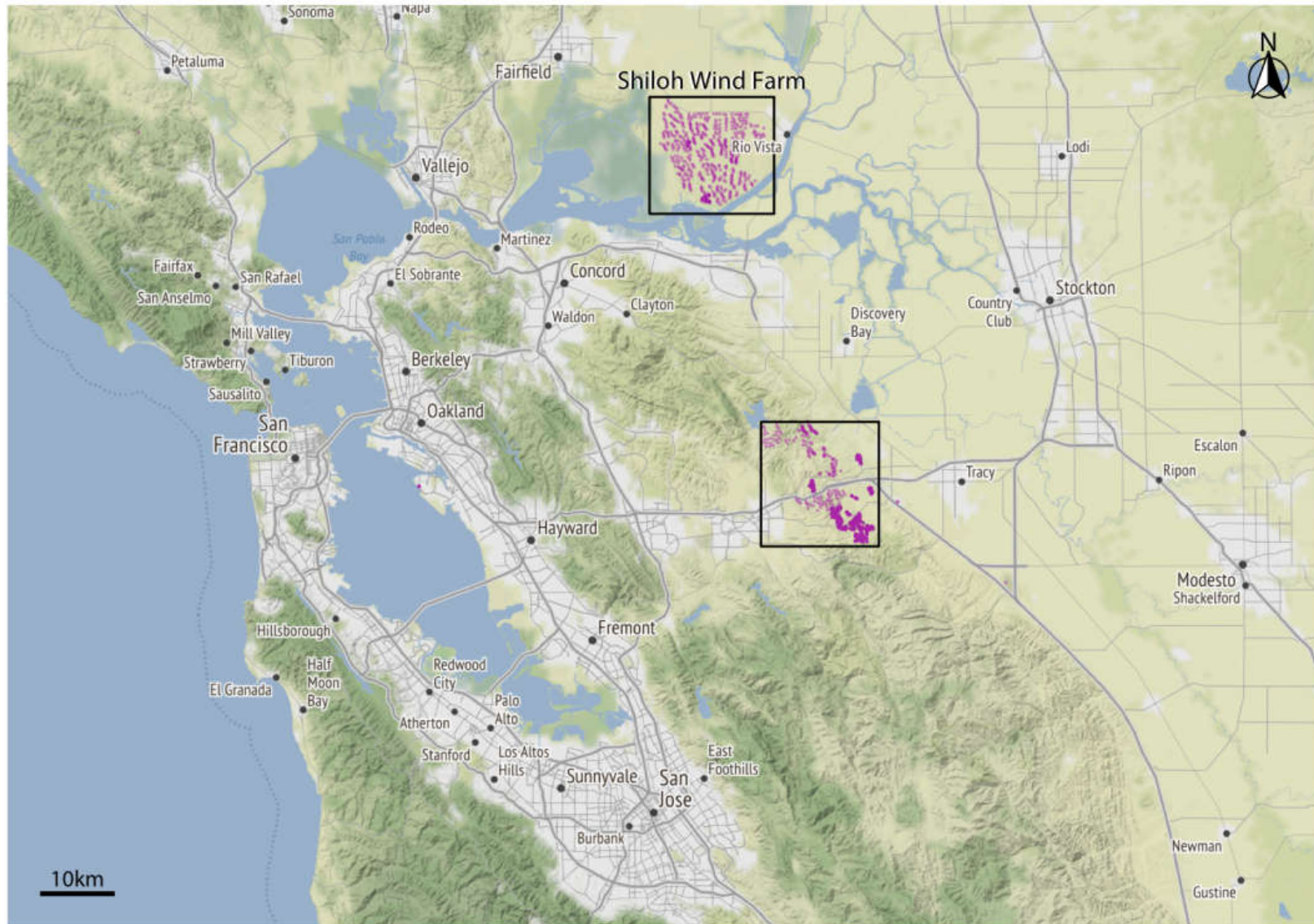
---

# VISUALIZING GEOSPATIAL DATA

- To see locations in the physical world
  - Map layers (corresponding projection)
  - Choropleth maps
  - Cartogram

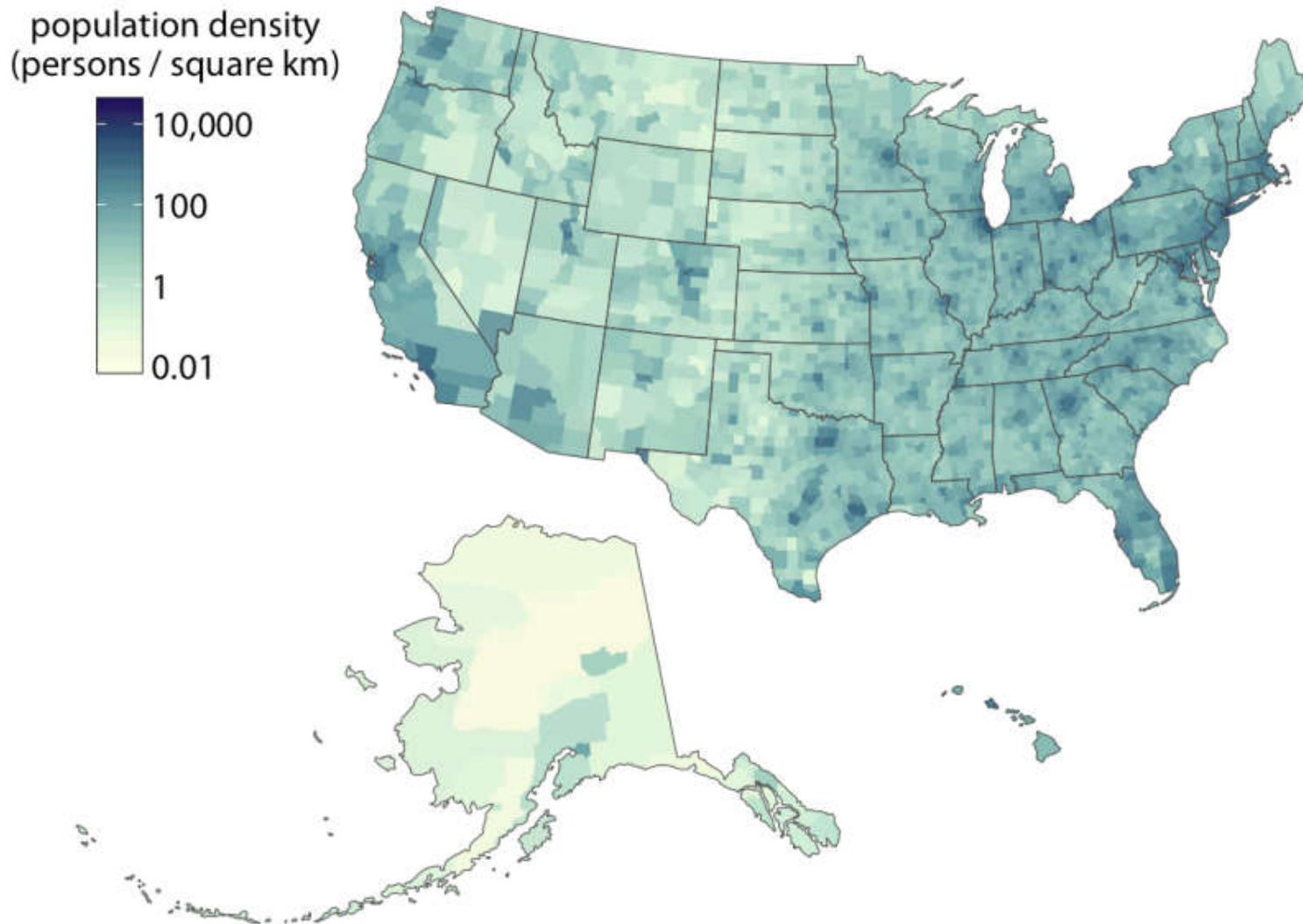


# MAP LAYERS



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

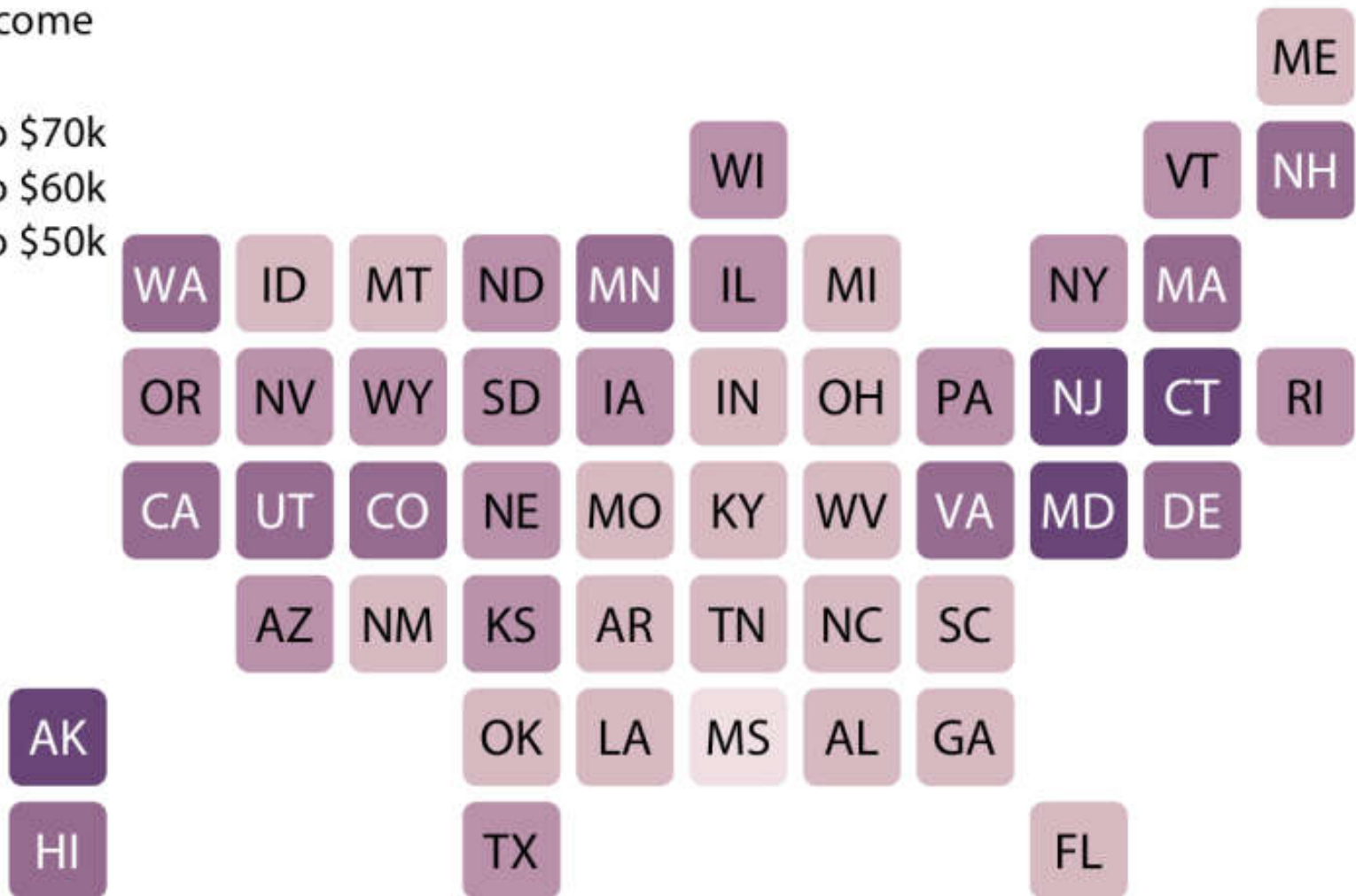
# CHOROPLETH MAPS



Claus O. Wilke, Fundamentals of data visualization: a primer on making informative and compelling figures, O'Reilly, 2019

# CARTOGRAM

median income

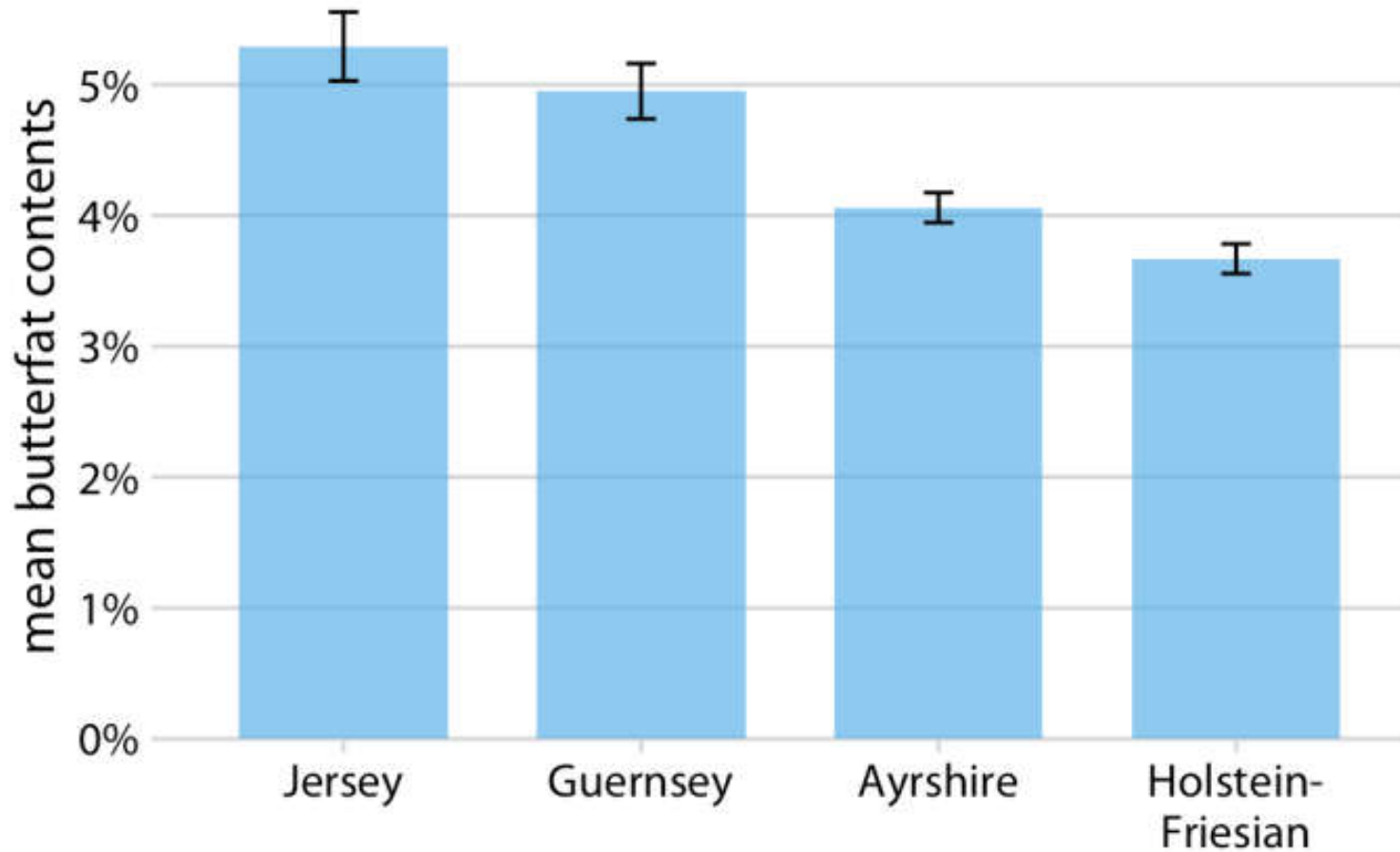


---

# VISUALIZING UNCERTAINTY

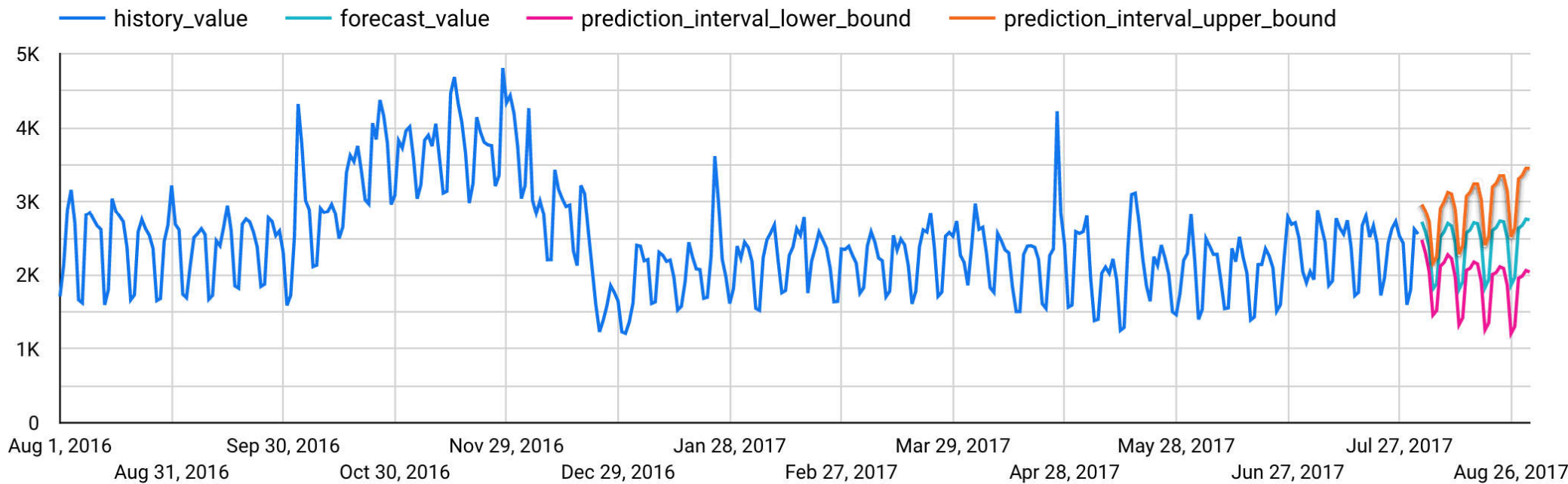
- Data with uncertainty
  - Error bars
  - Confidence bands

# ERROR BARS





# CONFIDENCE BANDS



<https://cloud.google.com/bigquery-ml/docs/arima-single-time-series-forecasting-tutorial>

---

# SUMMARY

- BI process and tasks
- Data visualization
  - Amount
  - Proportion
  - Distribution
  - Association
  - Time series
  - Trend
  - Geography
  - Uncertainty

# QUESTIONS AND ANSWERS



Picture from: <http://philadelphiasculpturegym.blogspot.com/2013/09/save-date-free-talk-and-q-on-affordable.html>



---

# REFERENCES

- [1] Tobias Zwingmann, AI-Powered Business Intelligence, Kindle Edition, O'reilly Press, 2022
- [2] Jiawei Han, Micheline Kamber, “Data Mining: Concepts and Techniques”, Third Edition, Morgan Kaufmann Publishers, 2012.
- [3] Jeen Su Lim, John Heinrichs, “Digital Business Intelligence Management with Big Data Analytics” Kindle Edition, O'reilly Press, 2021.
- [4] David L. Olson, Dursun Delen, “Advanced Data Mining Techniques”, Springer-Verlag, 2008.
- [5] Brian Larson, “Delivering Business Intelligence with Microsoft SQL Server 2016”, McGraw-Hill Education; 4 edition, 2016.
- [6] Oracle, “Data Mining Concepts”, 18c, E83730-03, 2018
- [7] Oracle, “Data Mining Application Developer’s Guide”, 2013.