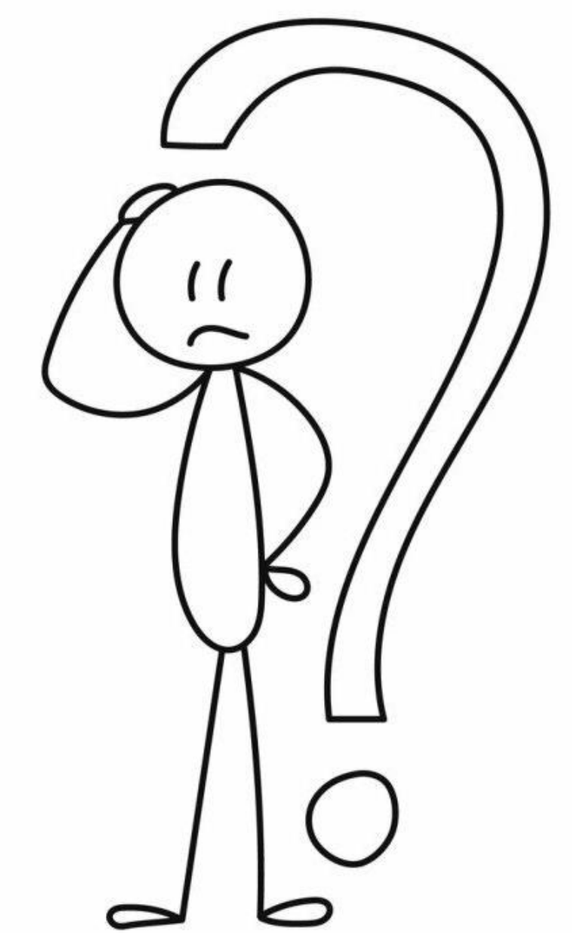
Visualization for Data Science
Bar Galore

Take Home Points

- To find meaning in what we see **we must selectively pay attention** to what is important
- Low-level vision is driven by object features rather than a conscious effort where to look (e.g., pre-attentive processing)
- Attention is driven by preexisting knowledge, expectations, and goals stored in long-term memory

Learning Outcomes

- Create the simple bar, column, sorted bar, grouped bar, stacked bar, and normalized bar charts
- Use aggregations and transformations prior to encoding data
- Customize various aspects of a chart
- Determine which bar chart variation is best suitable for a given task
- Describe the usefulness of the facet encoding channel



Prioritize choosing the most appropriate channel for each attribute

How do I pick *which* marks or channels to use?

Channel Characteristics

- Discriminability: how many unique steps can we perceive?
- Separability: is our ability to use this channel affected by another one?
- Popout: can things jump out using this channel?
- Grouping: can a channel show perceptual grouping of items?
- Accuracy: how precisely can we tell the difference between encoded items?

Channels: Rankings

Magnitude Channels: Ordered Attributes

- Position on common scale
- Position on unaligned scale
- Length (1D size)
- Tilt/angle
- Area (2D size)
- Depth (3D position)
- Color luminance
- Color saturation
- Curvature
- Volume (3D size)

Same

Same

Effectiveness

Best

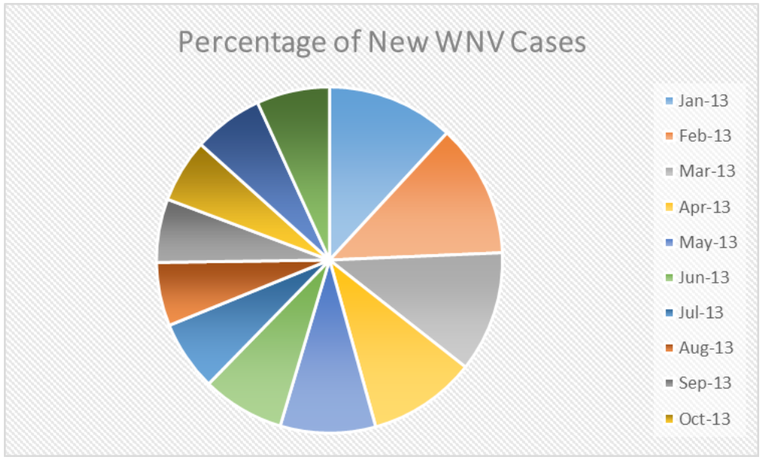
Least

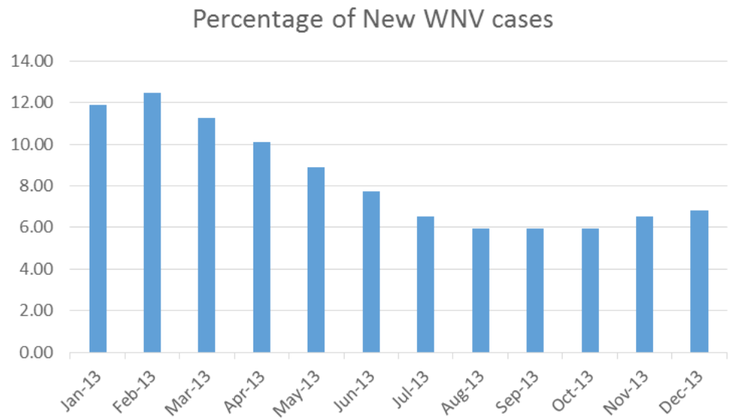
Identity Channels: Categorical Attributes

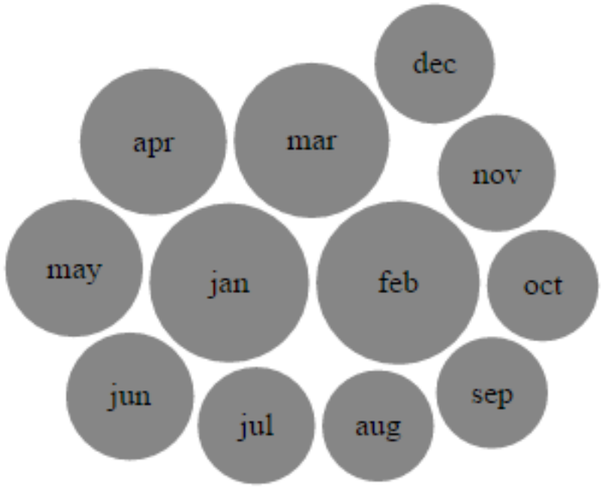
- Spatial region
- Color hue
- Motion
- Shape

- expressiveness
  - match channel and data characteristics
- effectiveness
  - channels differ in accuracy of perception
  - spatial position ranks high for both

Representation Effect



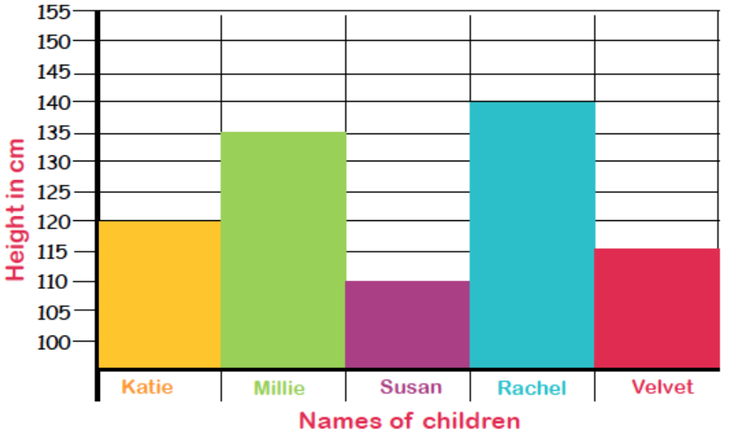


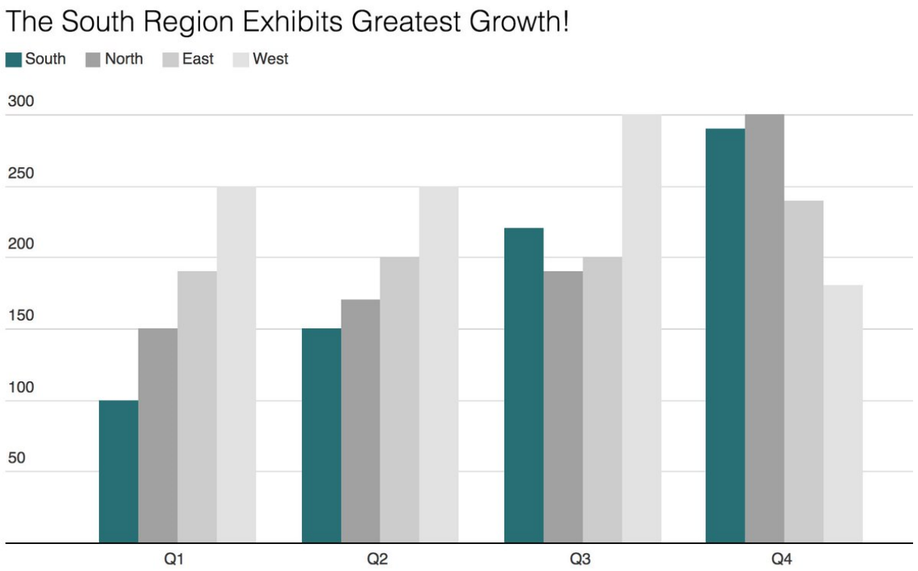


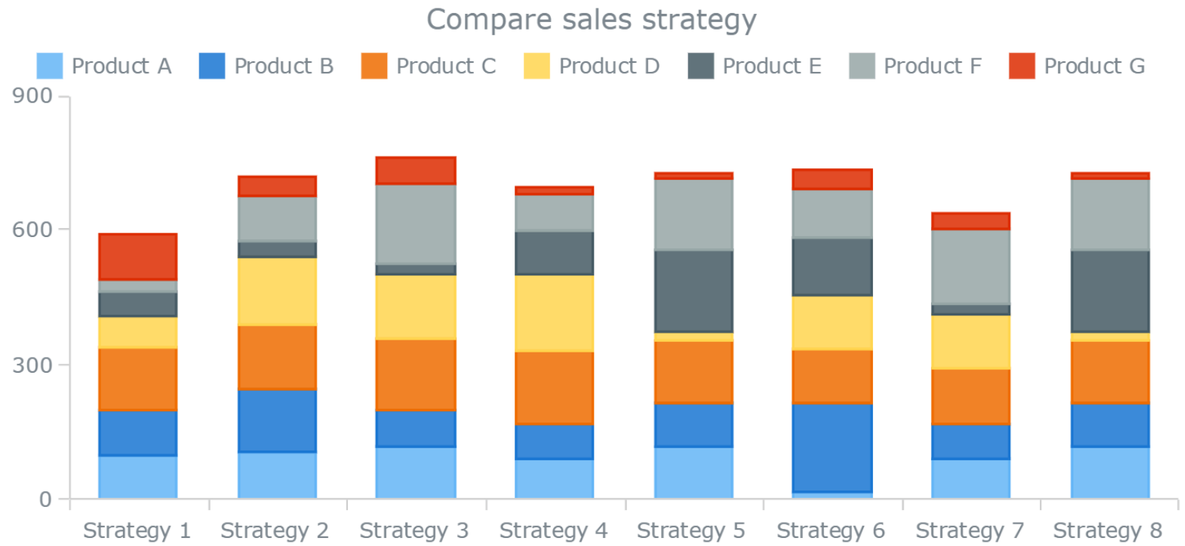
Clicker Question

Which visual cue do bar charts use to encode data

A. Position common scale
B. Position non-aligned scales
C. Length
D. Direction
E. Area



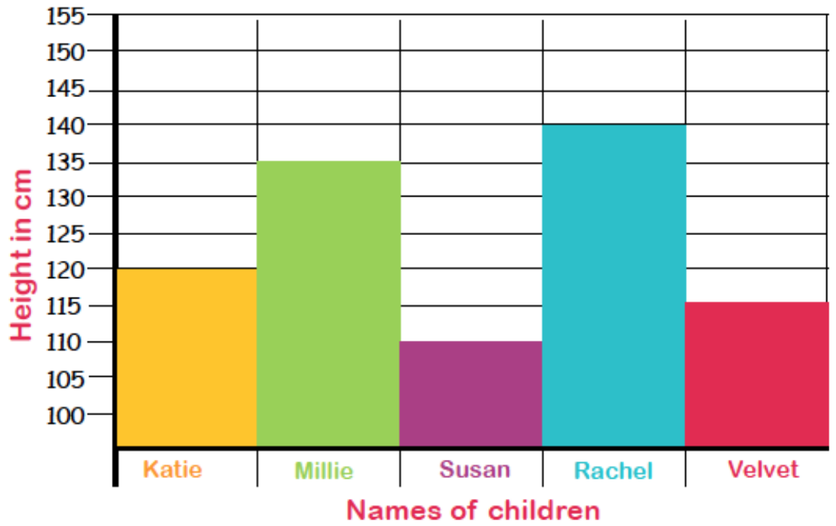




Bar Chart

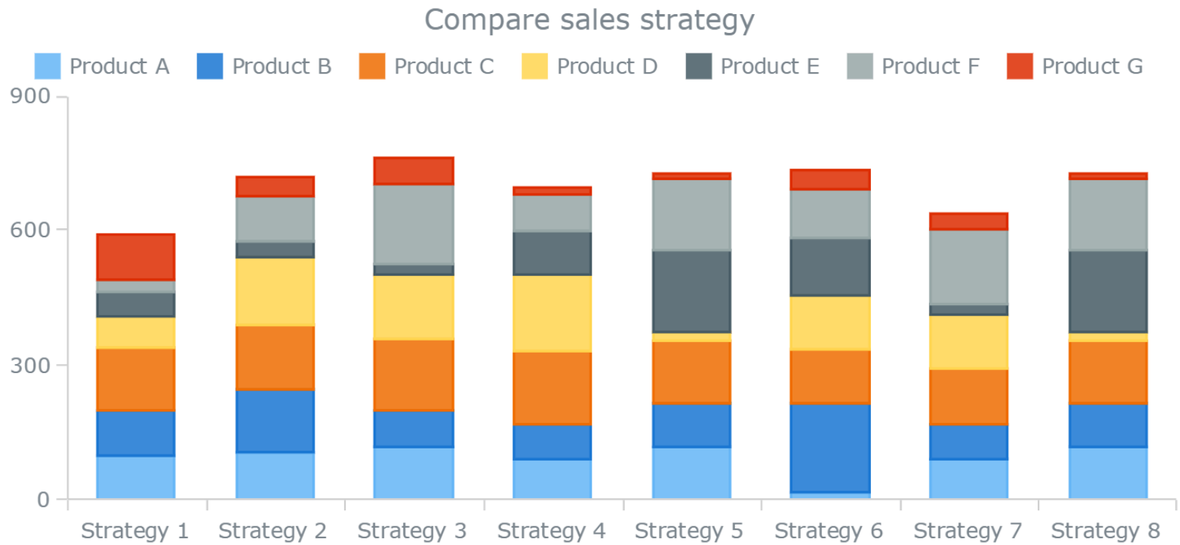
- Good for
- Comparisons

- Bar charts do not display **continuous developments** over an interval. Bar chart's discrete data is categorical data and therefore answers the question of "how many?" in each category.



Stacked Bar Chart

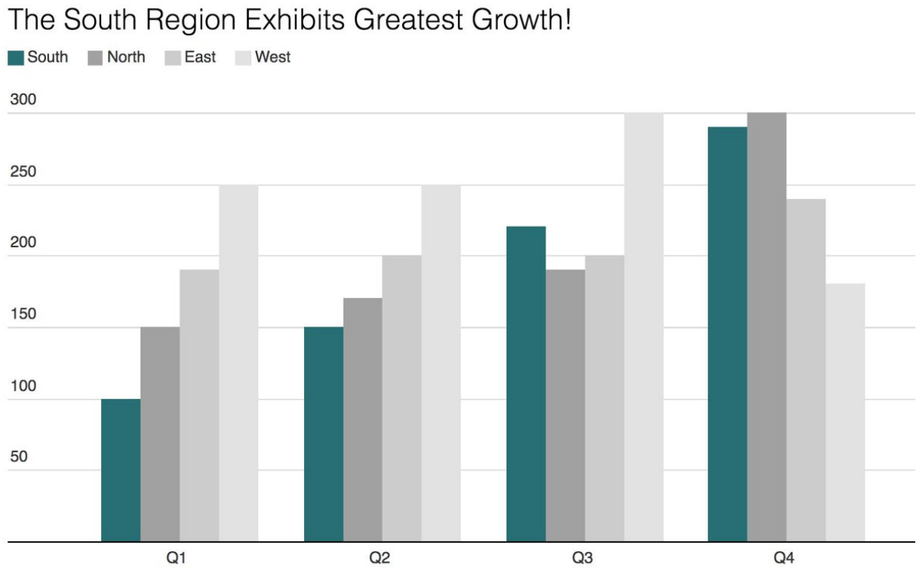
- Good for
- Proportions
- Parts to a whole
- Comparisons



- Limitation(s)
- Harder to read the more segments each bar has.
- Comparing segments to each other is difficult, as they're not aligned on a common baseline.

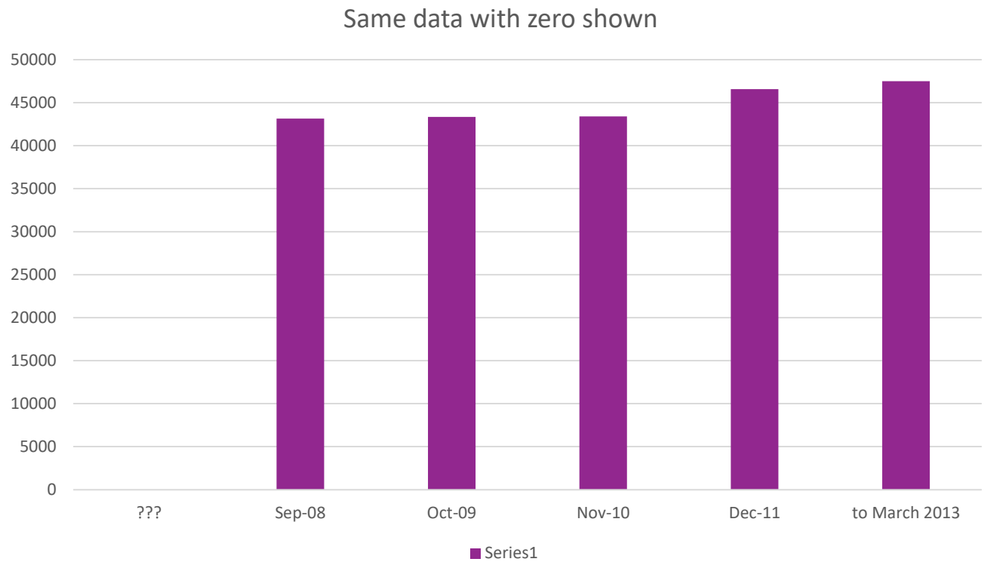
Multi-Set or Grouped Bar Chart

- Good for
- Distribution
- Relationships
- Comparisons



- The downside is that they become harder to read the more bars you have in one group.

Group Exercise: Is this visualization misleading? Why or why not?



Lab Groups – Upstairs RM 246

Last Name	Perferred Name
Lucas Bruxellas	Isabela
Mak	Kevin
Moynier	Lucas
Nguyen	Kaylie
Nguyen	Tien
Peng	Ryan
Pham	Chi
Purnomo	Annabelle
Radic	Sandra
Ren	Camilla
Reza	Naaimur
Si	MinZhang
Sun	Weihao
Thoma	Ethan
Wang	Minghao
Wang	Ziya
Wong	Gabrielle
Yang	Mina

Lab Groups – Downstairs RM 015

Acar	Berkay Talha
Ahn	Emily
BATTU	Riddhi
Chang	Benjamin
Chen	Adam
Chen	Jordon
Cho	Allan
De Ezkauriatza Sanchez	Javier
Delorme	Erika
Dorador	Emilio
Escolar Bach	Nitahi
Gillies	Matthew
Kong	Amy
Kung	Justin
Kwok	Silas
LE	Edison Le
Lee	Peter
Li	Davis

Lab

- Tomorrow is your second lab
- Make sure that the Setup is complete
- Make sure that you have done T1 – T5 prior to the lab
- The lab will be released at 11am.
- You must complete, get your tests to pass and then submit on Gradescope.
- TAs are their to support you, not to give you the answers
- During each lab add to your crib sheet, this is a list of common altair and pandas methods that you use, (will be helpful during the exam)
- Each lab is intended to be done individually, please do not share your work with anyone.
- There are 6 required labs, the other are just drop in office hours.

Lab Etiquette

- Lab begins at 11 or 12:30
- Students that arrive more than 5 minutes late will lose the etiquette marks for the day. Arriving late requires the TAs to repeat their summaries of what the lab entails which has a negative impact on the learning environment.
- Etiquette points include
  - Arriving on time
  - Working actively on the lab for an hour
  - Not engaging in academic misconduct
  - Represents 10% of your lab grade.

Office hours start next week

- TA Office Hours
  - Tuesdays in RM 238 at 5pm
  - Wednesdays Online on Zoom at 5pm
  - Saturday Online on Zoom at 2pm
  - Sunday Online on EdStem at 4pm
- Instructor Office Hours
  - In Person RM 238 from 1 – 3pm on Wednesdays

Next on Viz

- Lab 2 due in Lab tomorrow
- Assignment 2 out today due next Wednesday Feb 1st