Meetup 2: Graphical Perception Theory

George Hagstrom

Why Are Pie-Charts Terrible?

- "Pie Charts are Evil"
  - Chapter 2 Storytelling with Data

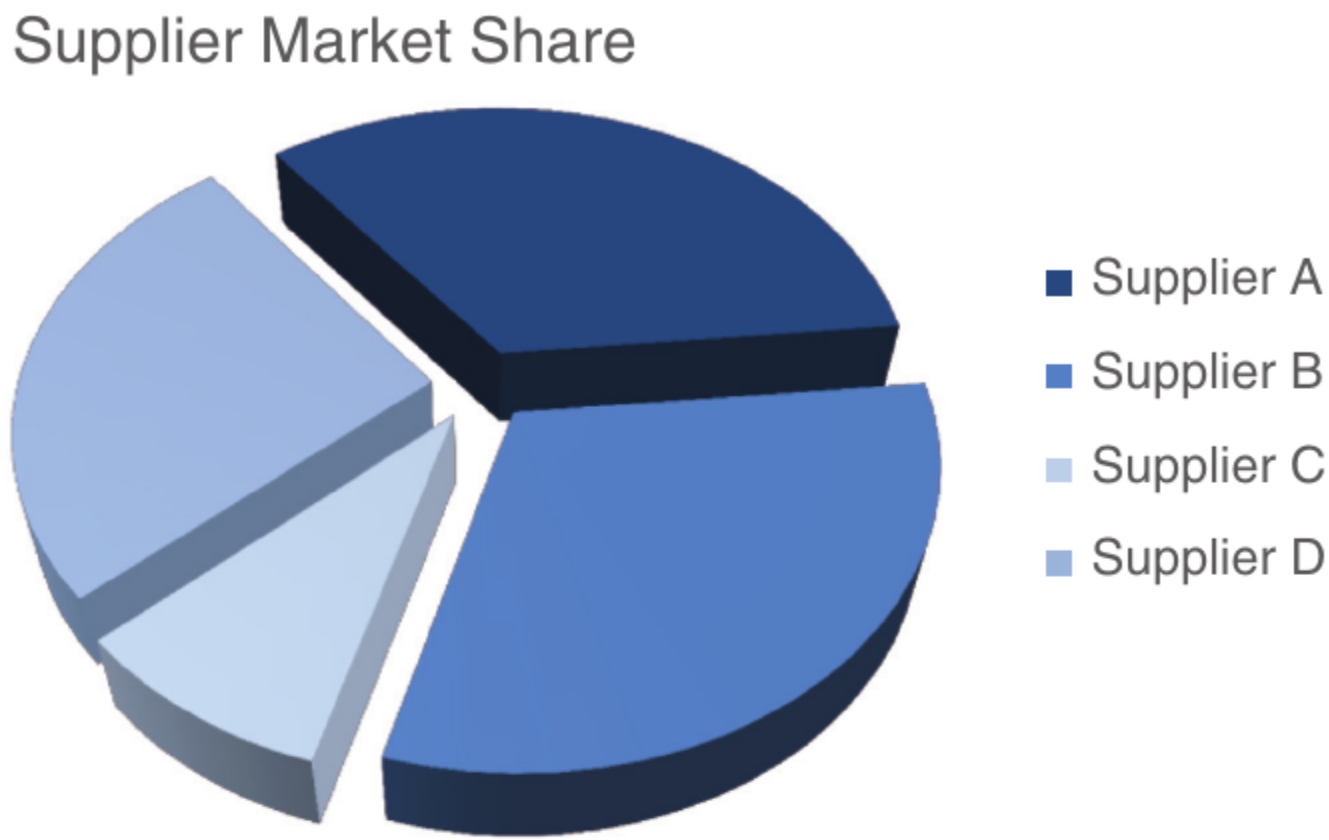


FIGURE 2.21 Pie chart

Why Are Pie-Charts Terrible?

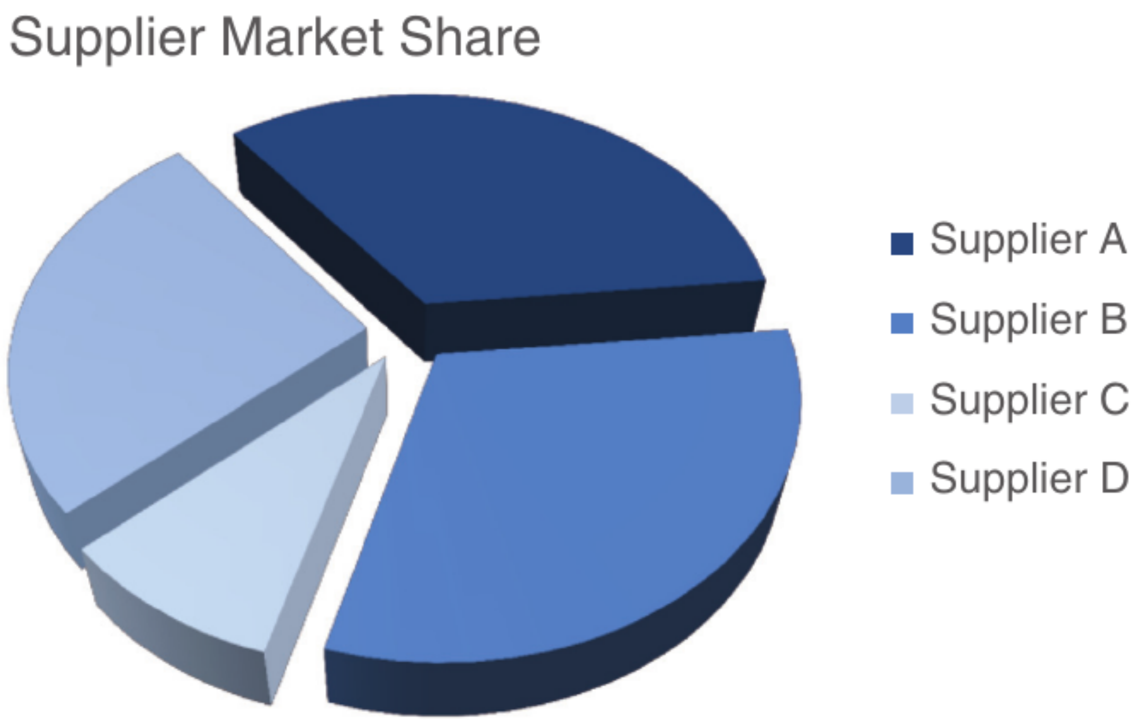


FIGURE 2.21 Pie chart

- Lots of graphics advice is presented as rules
- Are there principles behind these rules?

What to do this week

1. Readings: Chapter 2 of "Storytelling" and Chapter 5 of "Fundamentals"
2. Bonus Reading: Cleveland et al 1984 Perceptual Visualization Paper
3. Discussion: Favorite Graph
4. First HW due Sunday at midnight

What Makes a Graph Good

- We use graphs to communicate quantitative data
- A graph is good if it successfully conveys quantitative data to the reader
- Can we break down visualizations into a series of elementary perceptual tasks and then test how good humans are at completing each one?
- Research program led by many, William Cleveland most prominent

Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. Angle
6. Area
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation

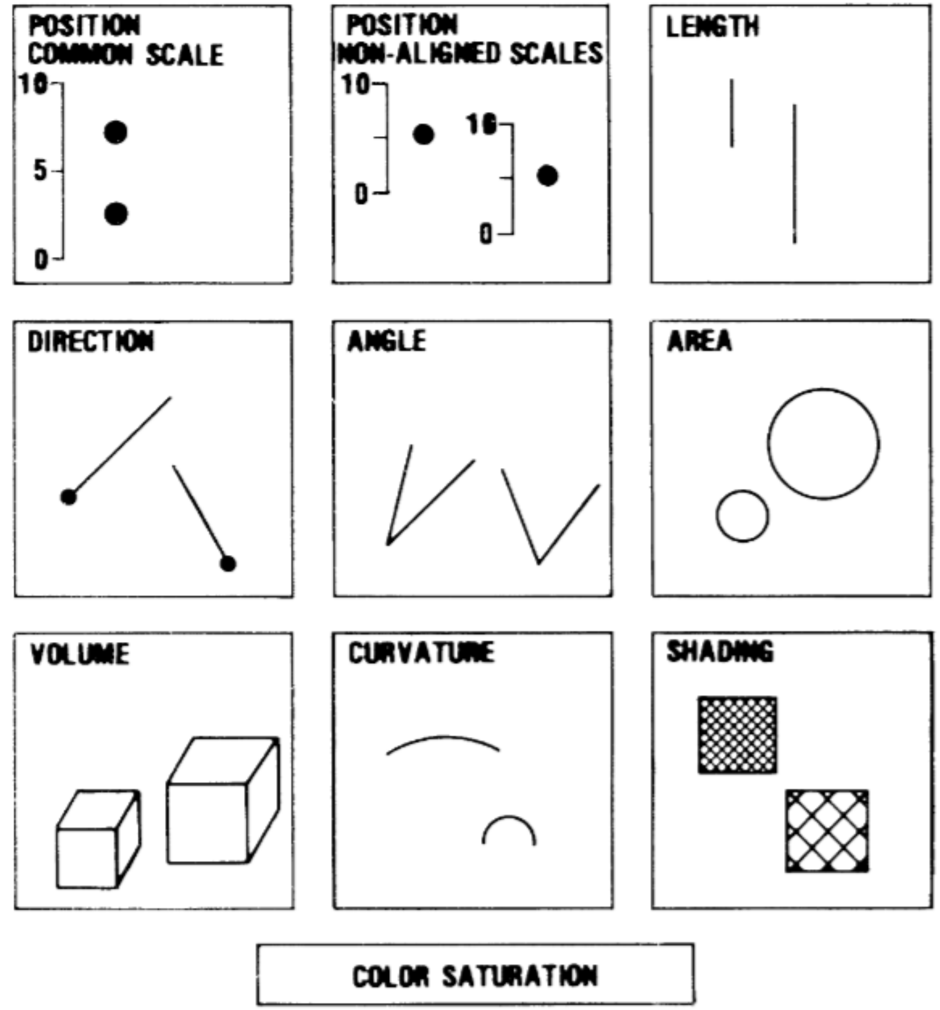


Figure 1. Elementary perceptual tasks.

Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. Angle
6. Area
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation

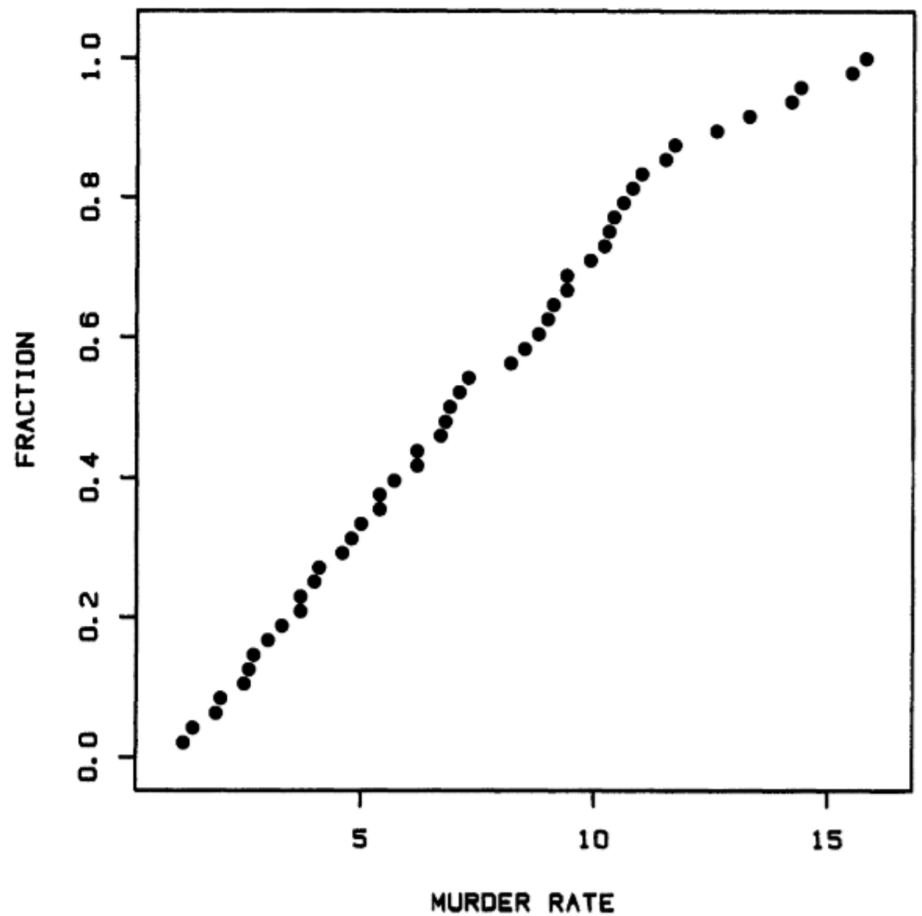
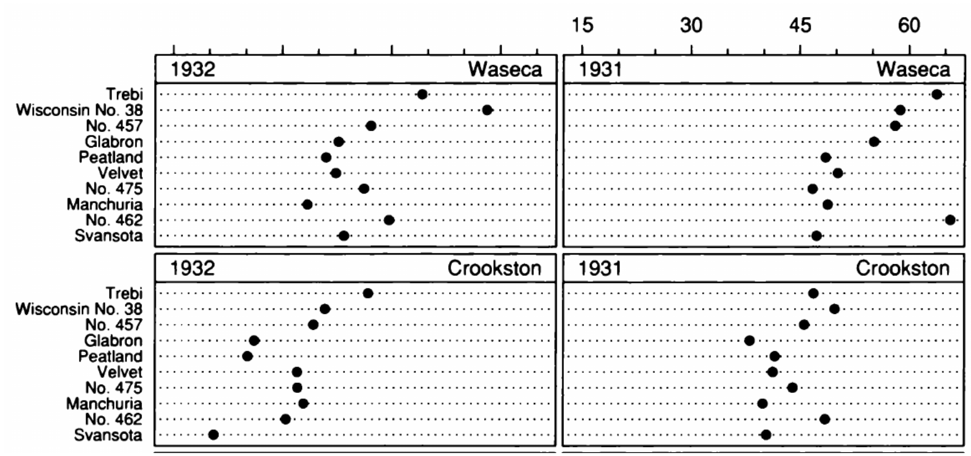


Figure 2. Sample distribution function of 1978 murder rate.

CUNY School of Professional Studies

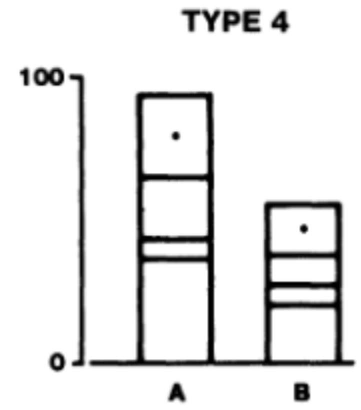
Elementary Perceptual Tasks

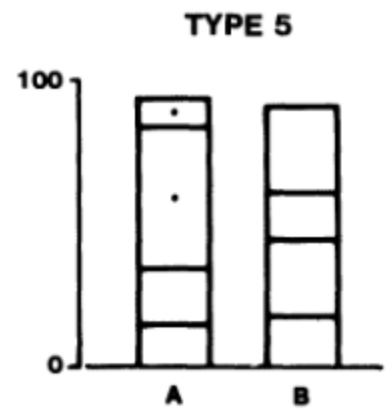
1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. Angle
6. Area
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation



Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. Angle
6. **Area**
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation





Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. **Length**
4. Direction
5. Angle
6. Area
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation

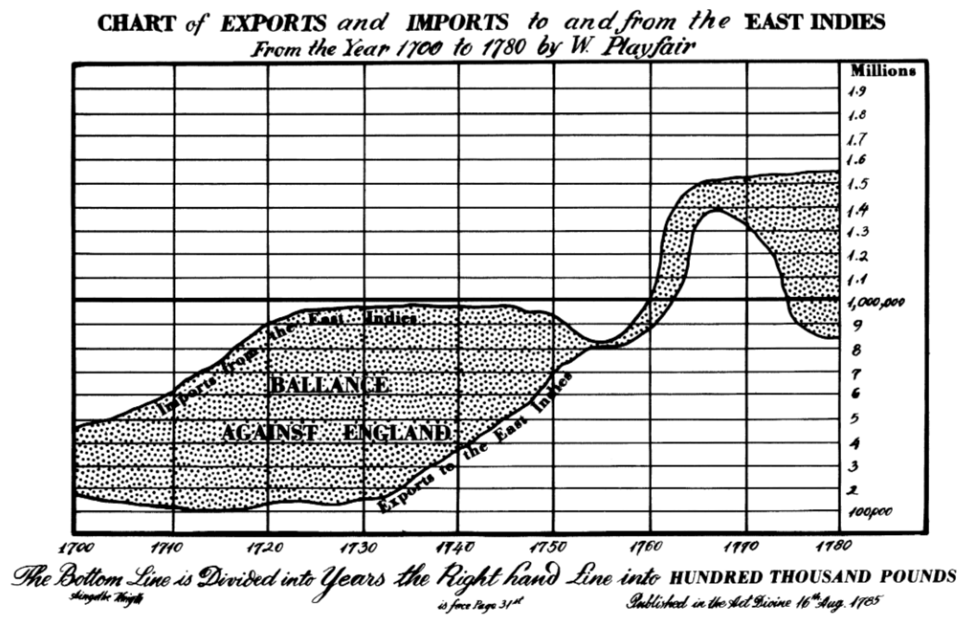
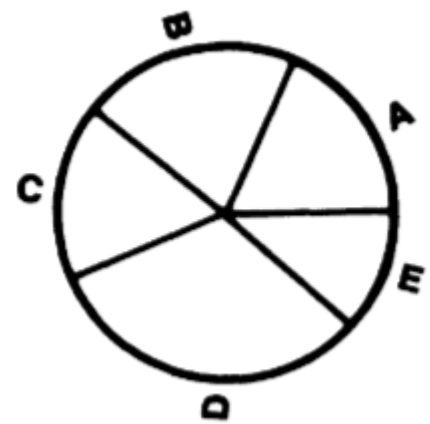


Figure 6. Curve-difference chart after Playfair.

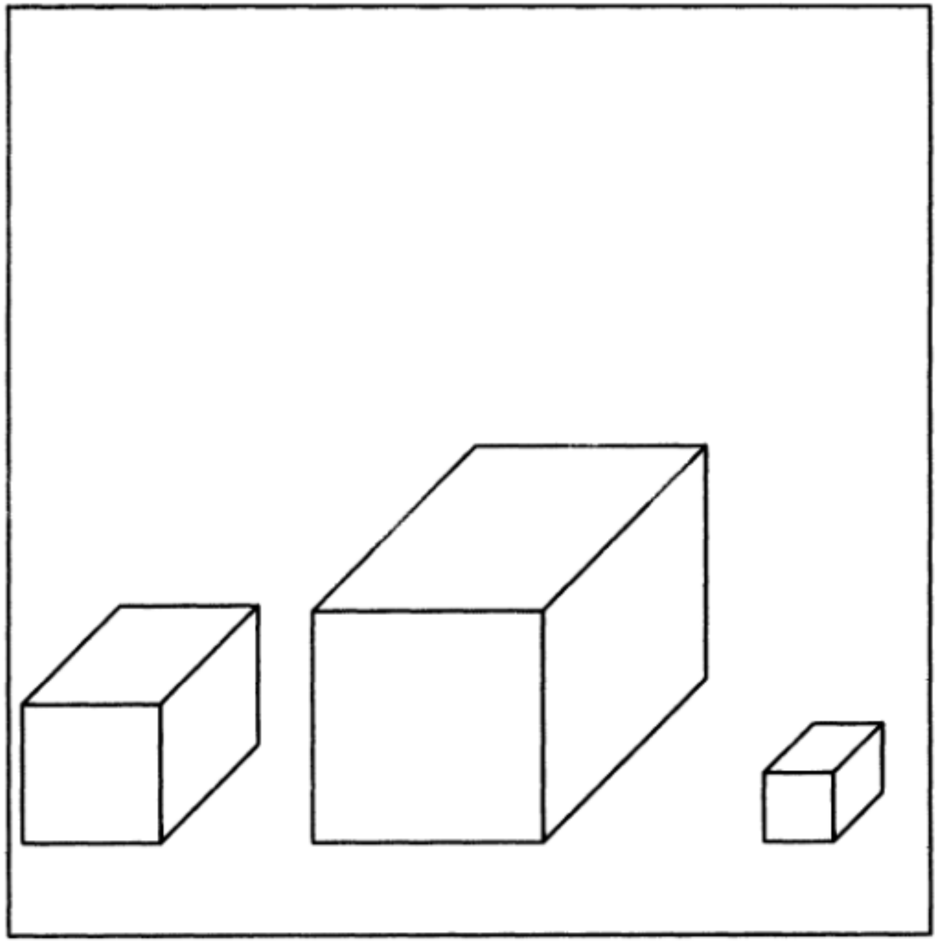
Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. **Angle**
6. **Area**
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation



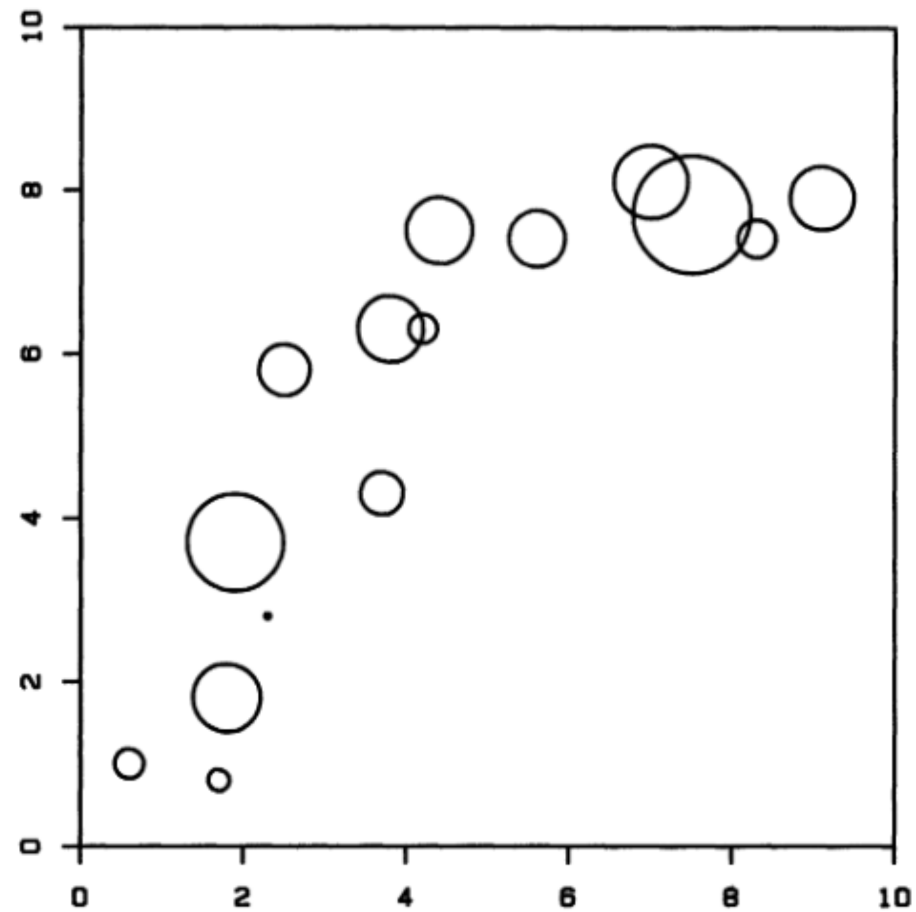
Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. Angle
6. Area
7. **Volume**
8. Curvature
9. Shading
10. Hue/Saturation



Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. **Length**
4. Direction
5. Angle
6. **Area**
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation



Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. Angle
6. Area
7. Volume
8. Curvature
9. **Shading**
10. Hue/Saturation

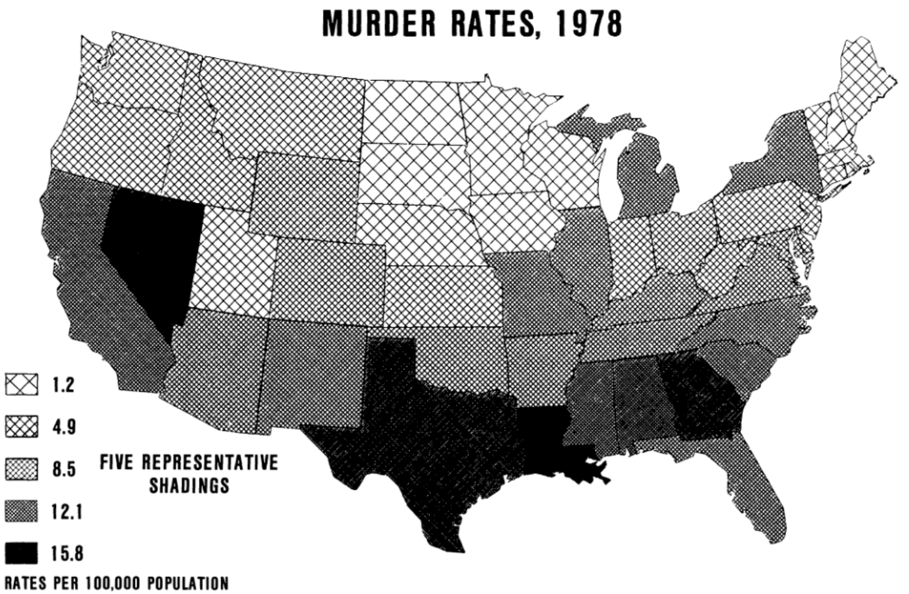
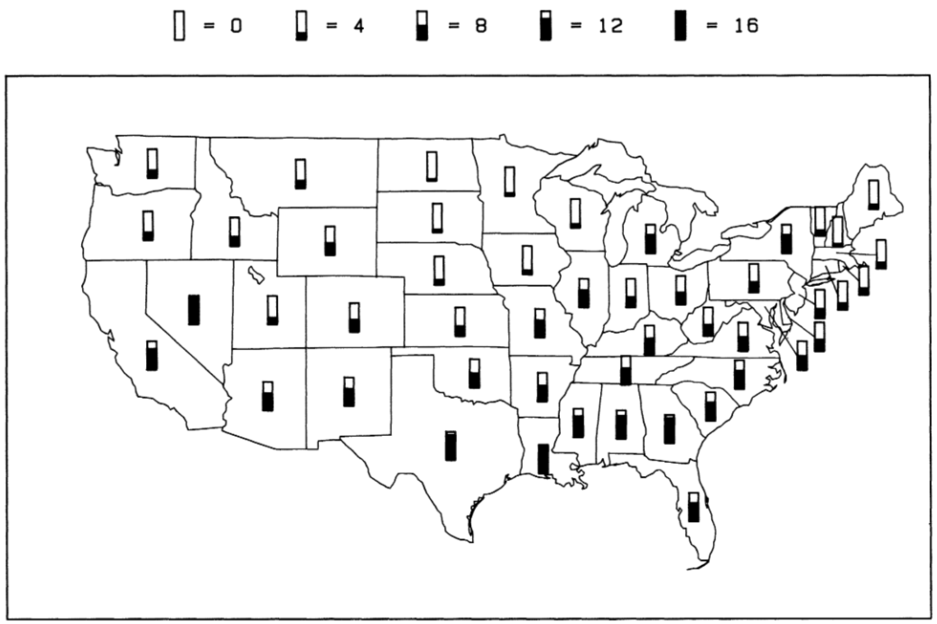


Figure 5. Statistical map with shading.

CUNY School of Professional Studies

Elementary Perceptual Tasks

1. Position along a common scale
2. Position along non-aligned scales
3. **Length**
4. Direction
5. Angle
6. **Area**
7. Volume
8. Curvature
9. Shading
10. Hue/Saturation

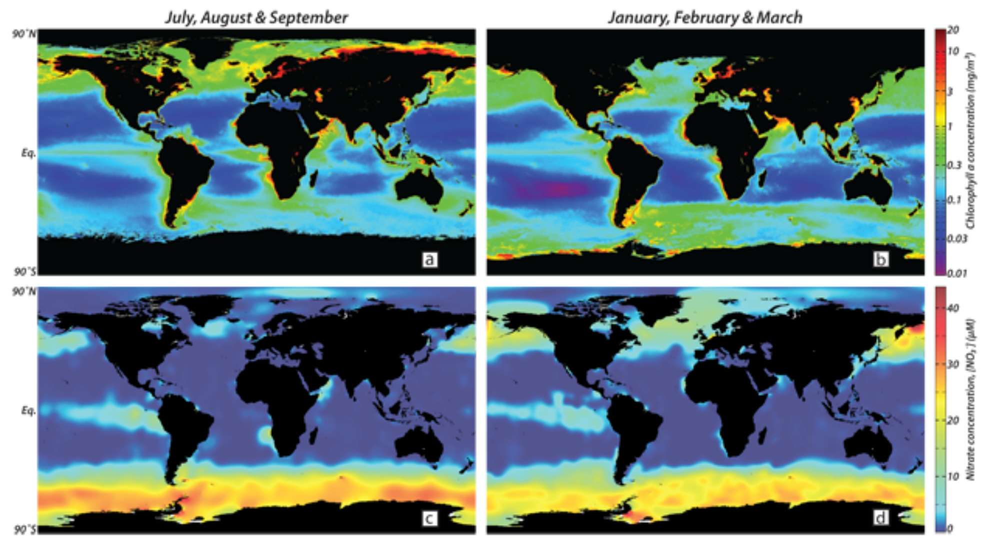


MURDER RATES PER 100,000 POPULATION, 1978

CUNY School of Professional Studies

Elementary Perceptual Tasks

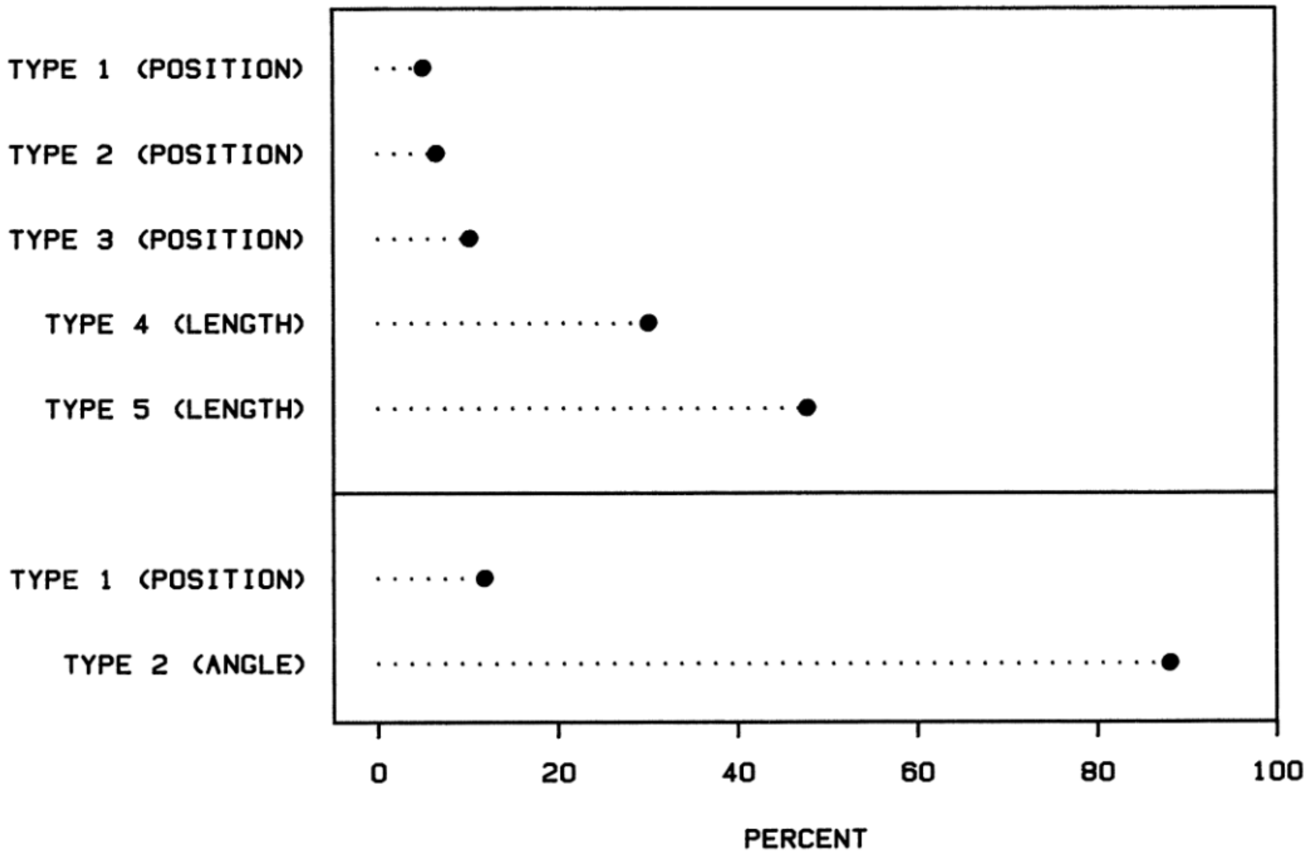
1. Position along a common scale
2. Position along non-aligned scales
3. Length
4. Direction
5. Angle
6. Area
7. Volume
8. Curvature
9. Shading
10. **Hue/Saturation**



CUNY School of Professional Studies

Some Perceptual Tasks are Easier

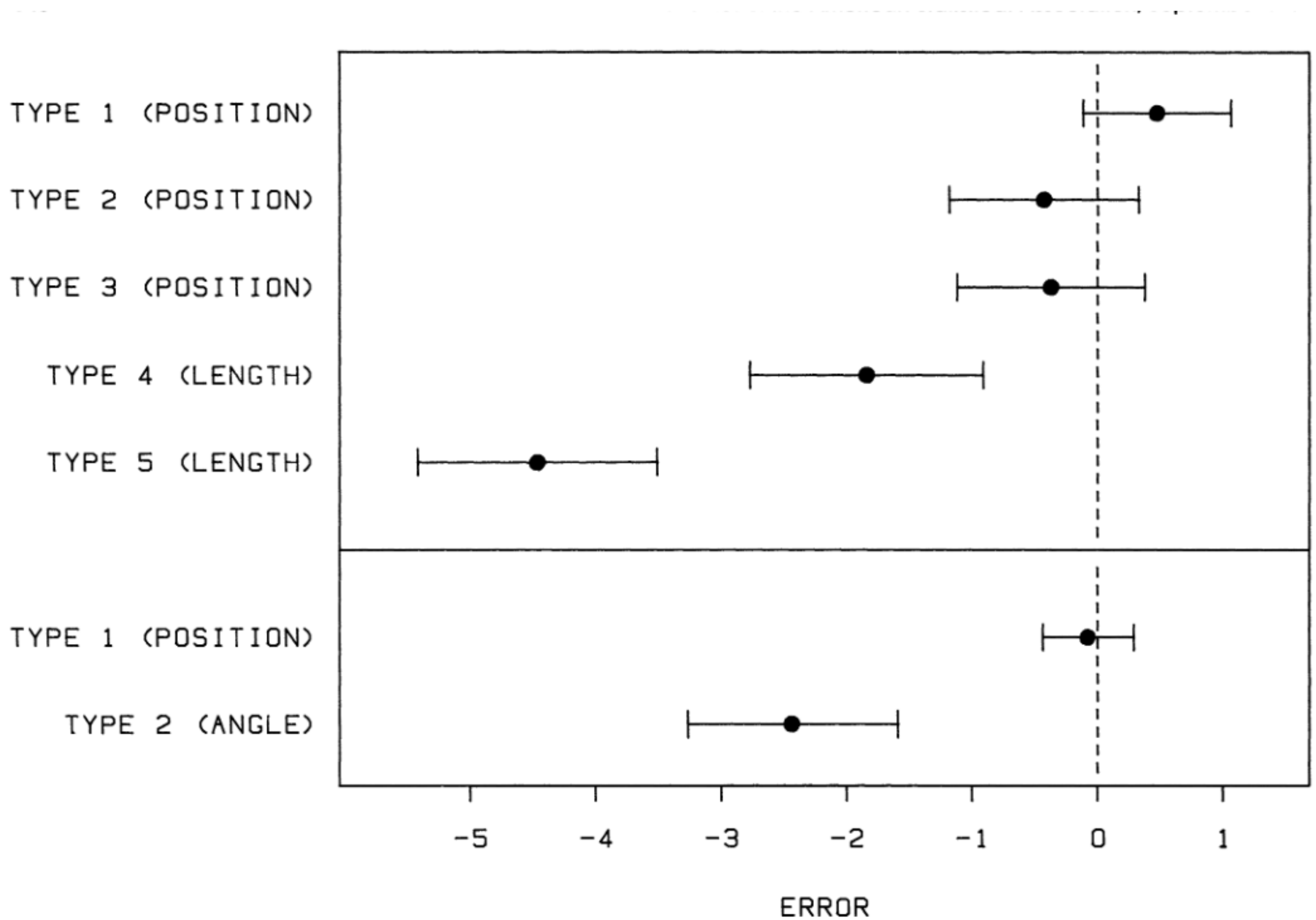
- Position along a common axis is the best
- Percentage of major errors:



CUNY School of Professional Studies

Some Perceptual Tasks are Easier

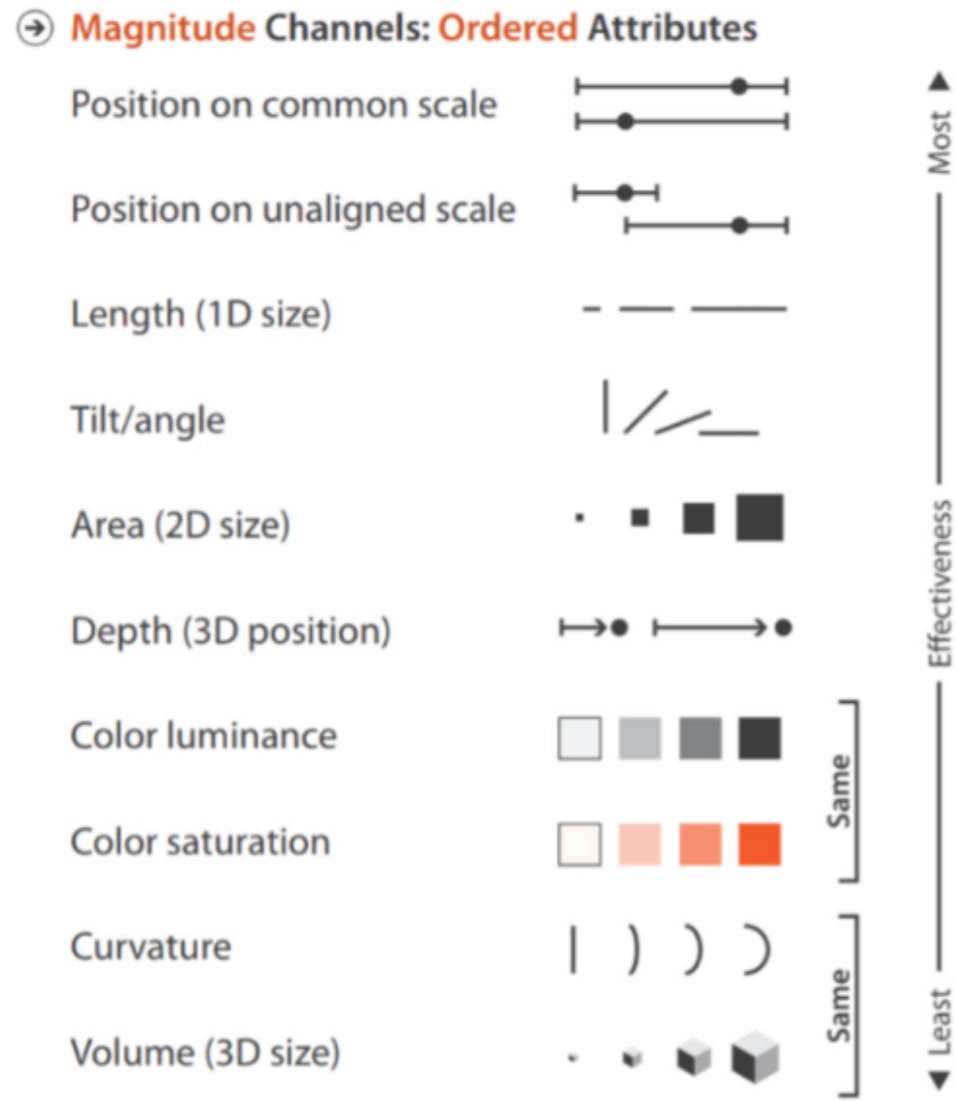
- Position along a common axis is the best



- Error

Overall Ranking

1. Position along a common axis
2. Position along an uncommon axis
3. Length, Direction, Angle
4. Area
5. Shading, Color Saturation
6. Volume, Curvature



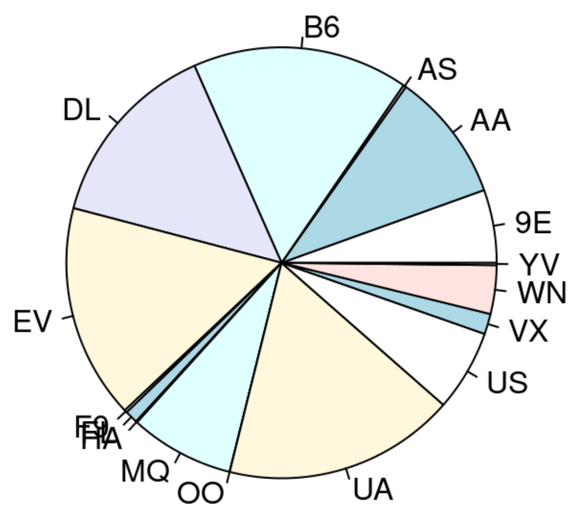
Munzner 2014

Implications

- Plot the comparisons you want to make on the same axis
- Position can usually substitute length, direction, and angle
- Avoid 3D
- Be cautious with area
- Shading and color saturation are very imprecise

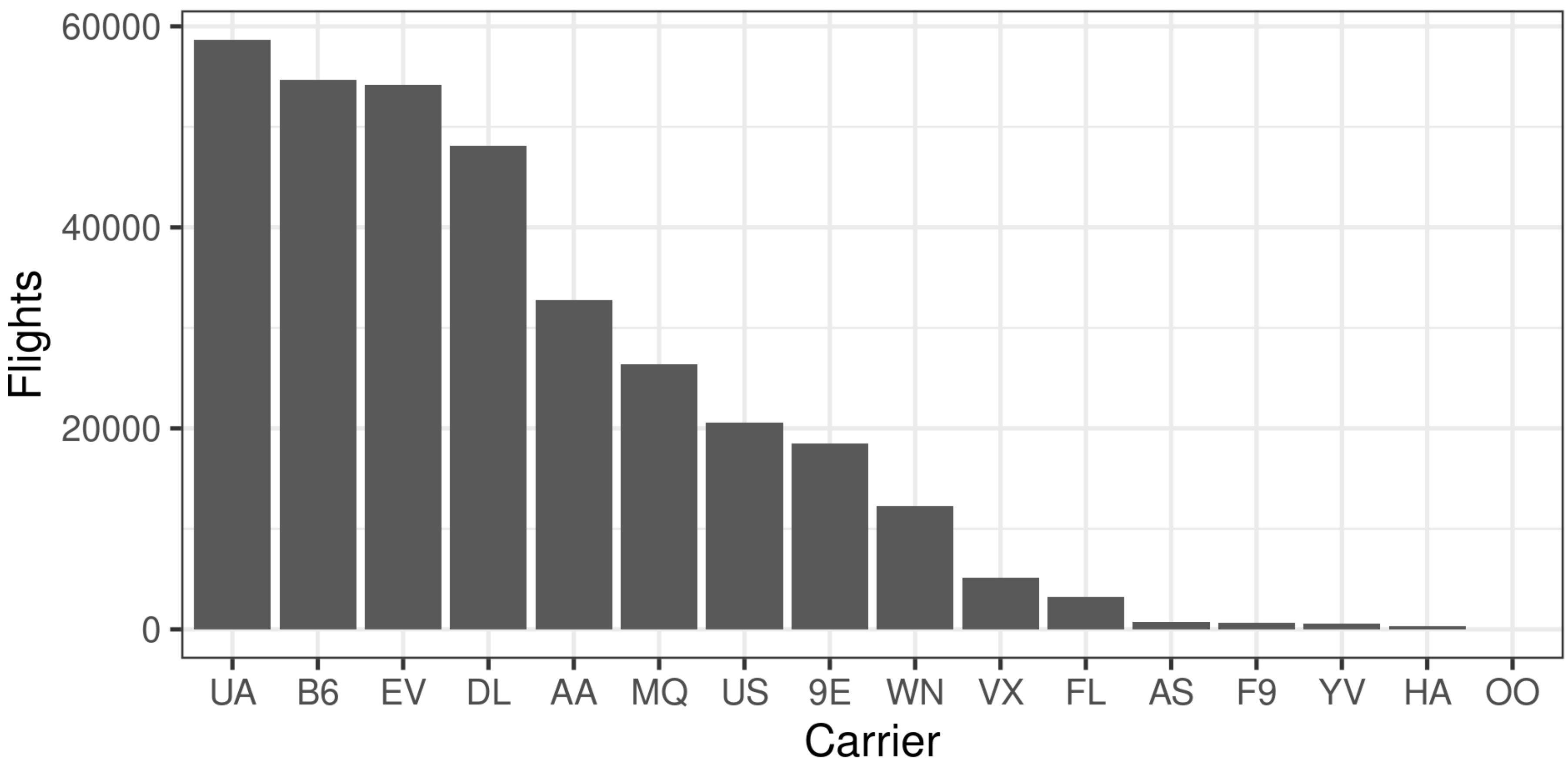
Dot-Plots for 1D Comparisons

- Flights from each carrier leaving NYC in 2013



Dot-Plots for Comparisons

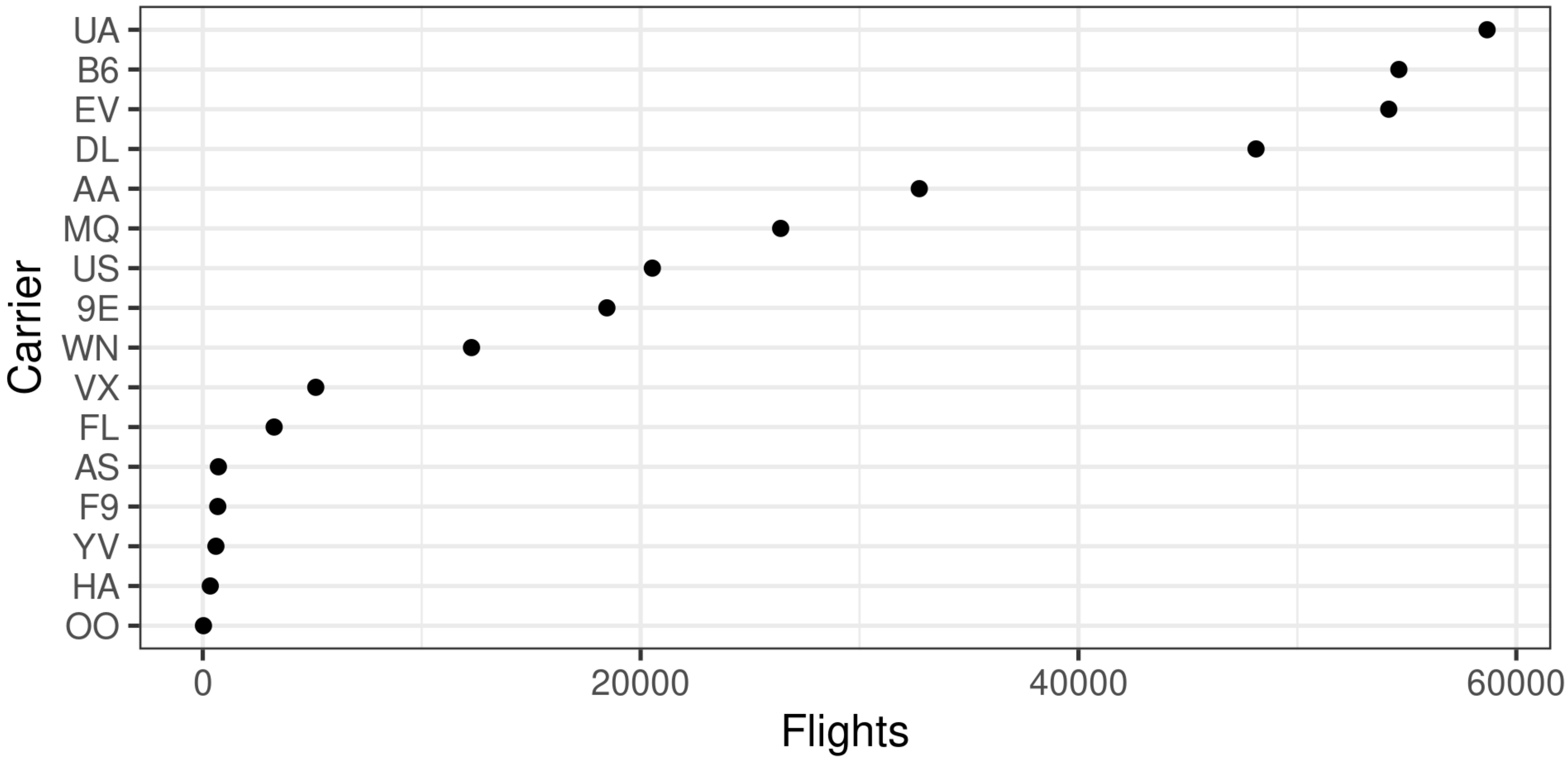
- Flights from each carrier leaving NYC in 2013



CUNY School of Professional Studies

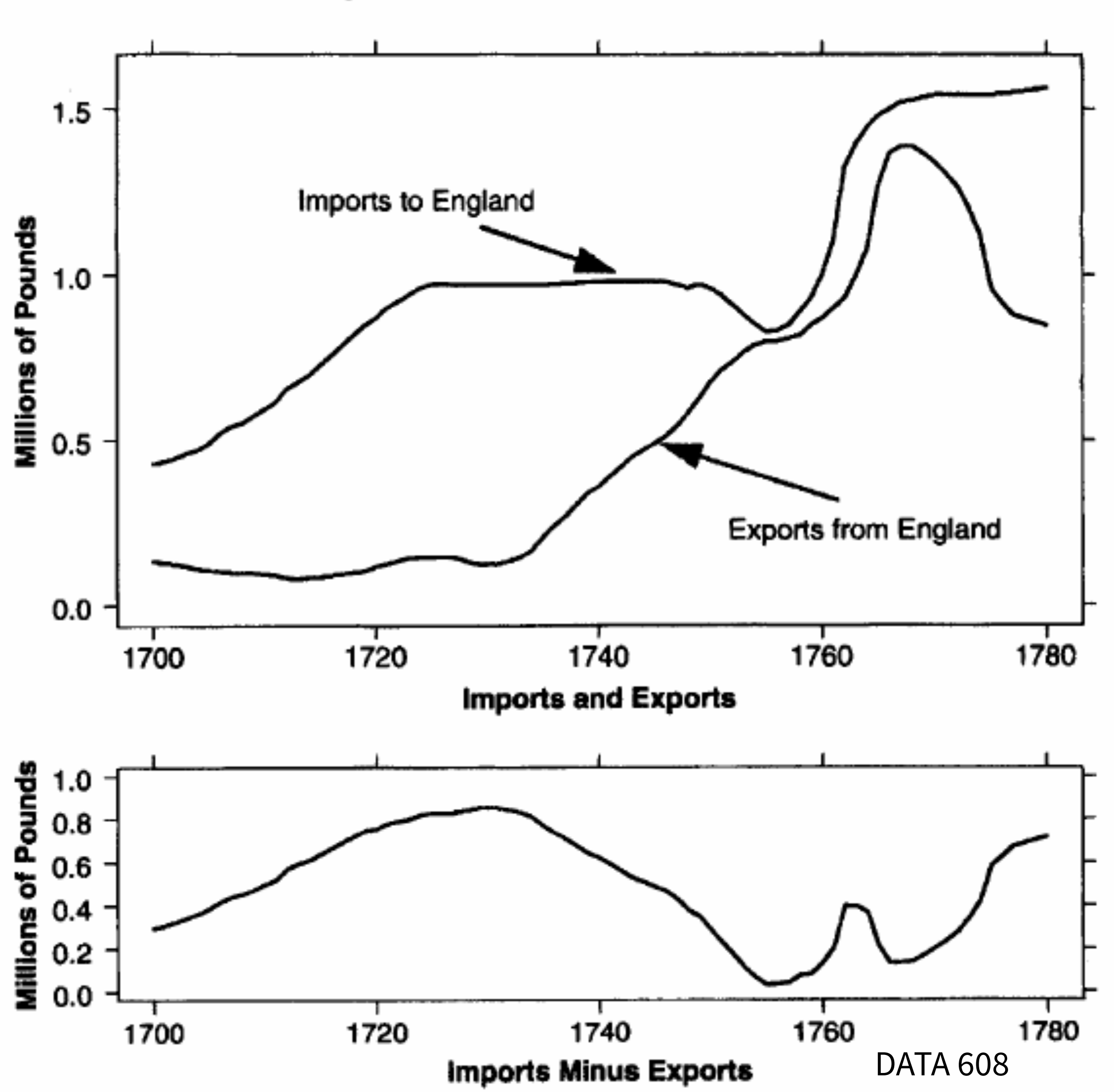
Dot-Plots for Comparisons

- Flights from each carrier leaving NYC in 2013



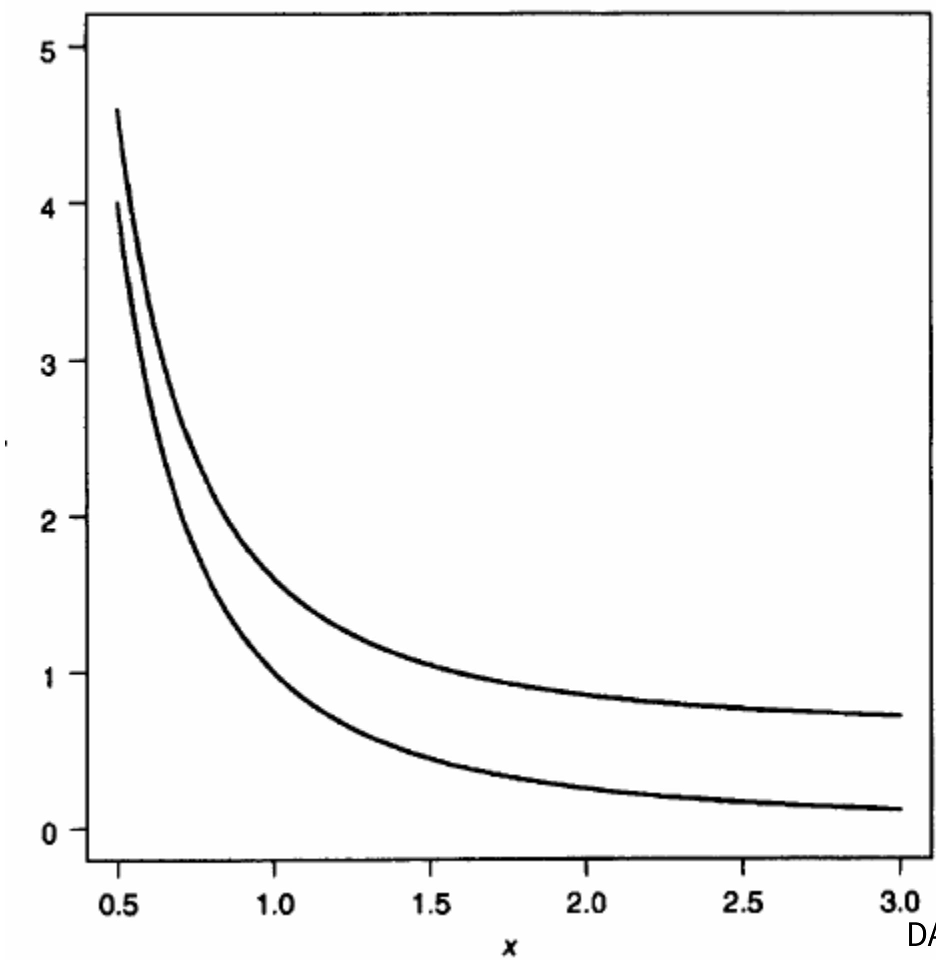
Show Differences Directly

- Position is much easier to judge than length



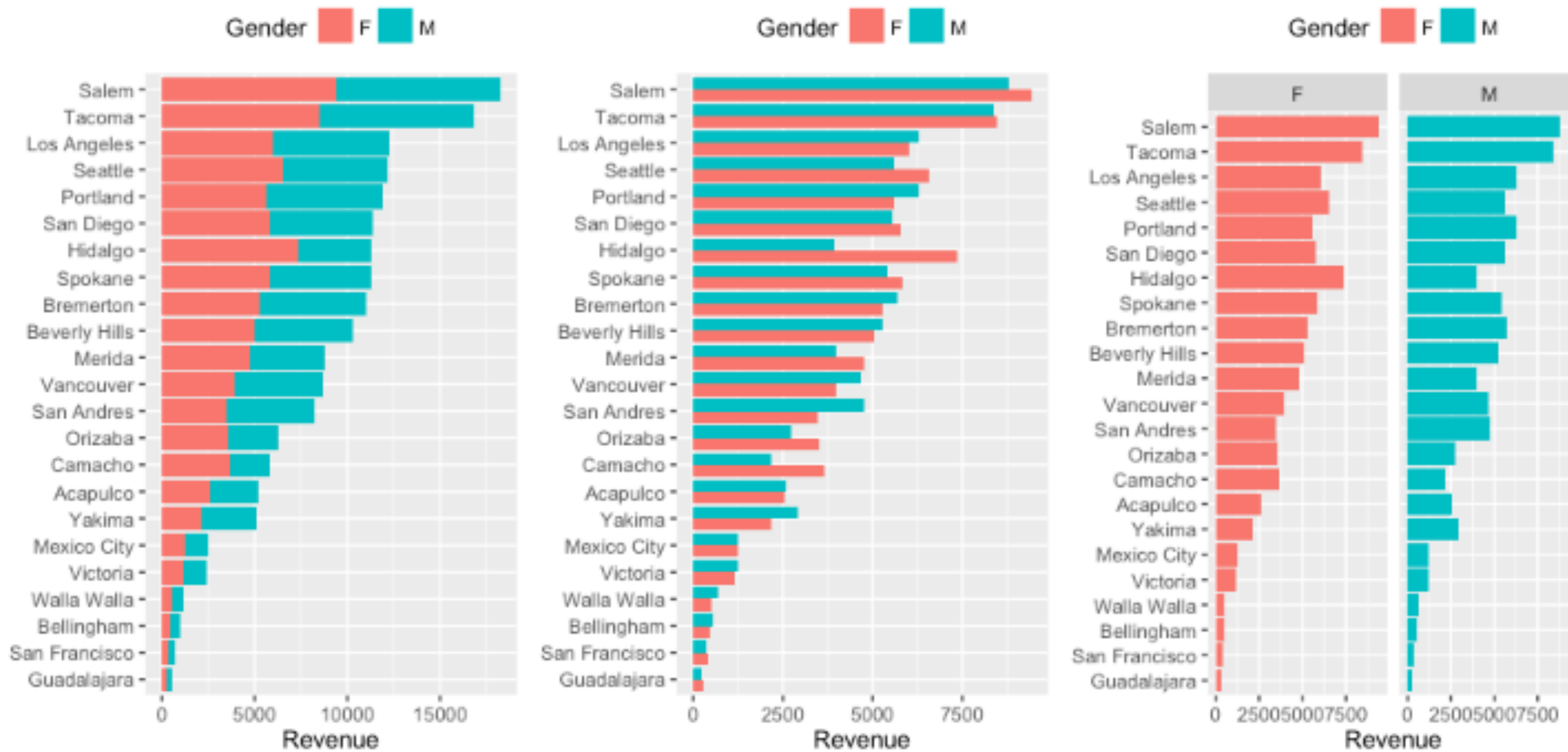
Show Differences Directly

- Position is much easier to judge than length
- These curves have a constant height difference



Show Differences Directly

- Adjacent Bars Versus Dots on the same Axis

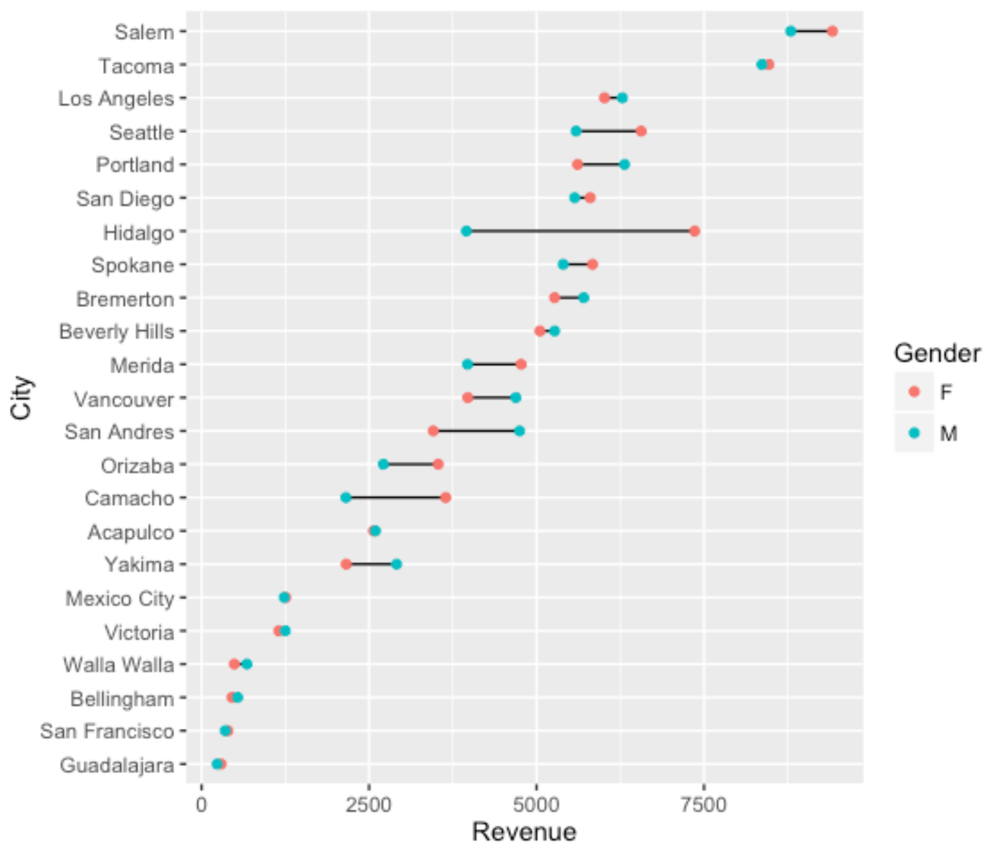


UCR Bus. Analytics

CUNY School of Professional Studies

Show Differences Directly

- Adjacent Bars Versus Dots on the same Axis



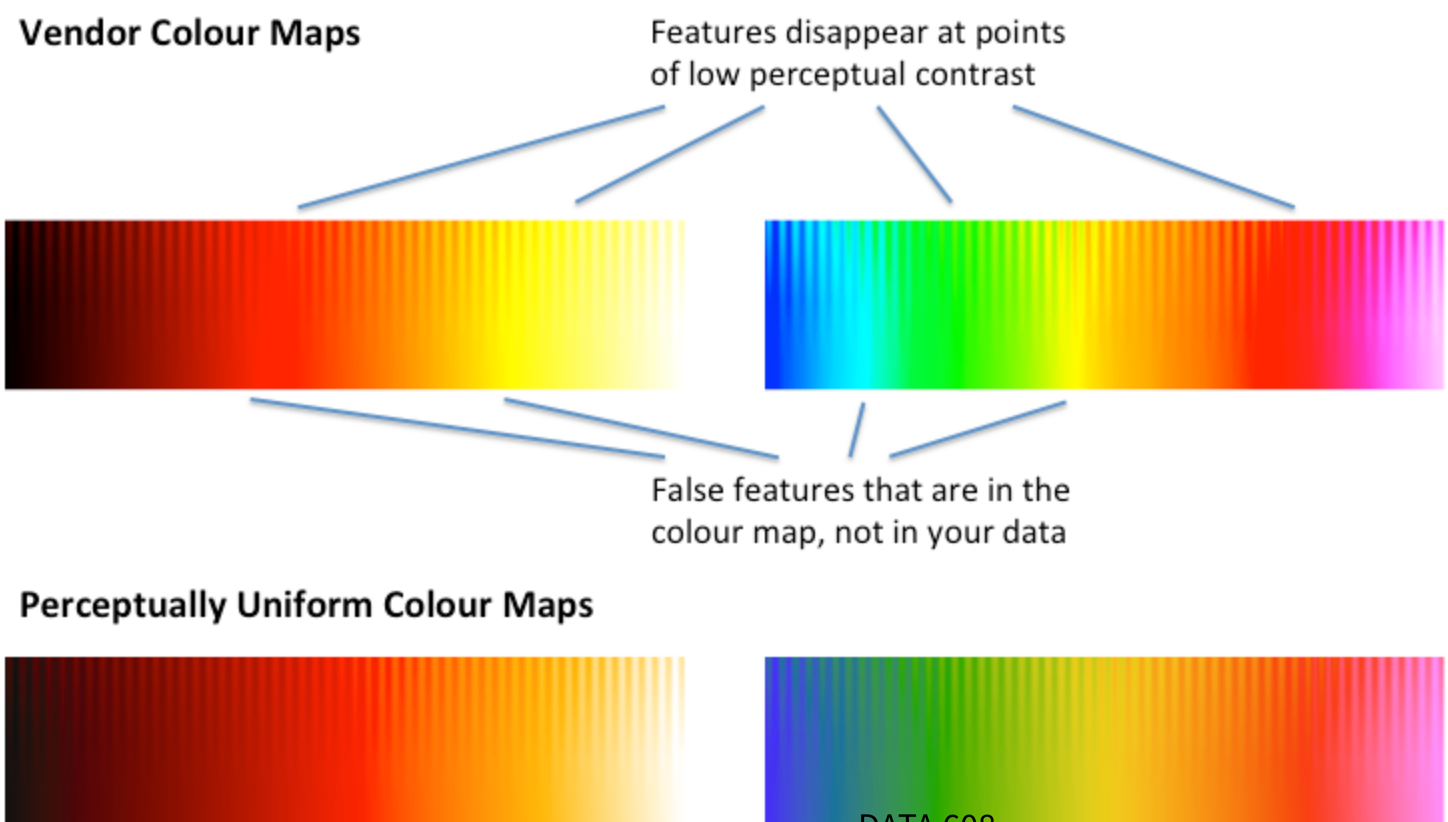
UCR Bus. Analytics

What to do About Color?

- Color is a poor choice for precise numerical comparisons
- Consider whether you need to use it at all
- It is very helpful for illustrating geographic patterns

Perceptually Uniform Colormaps

- Colormaps attach color to number
- Visual complexity makes it easy for the colormap to create features that aren't there or to hide features

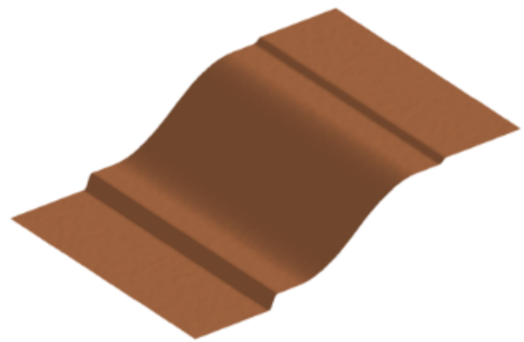


Colormap Example

- Which colormap renders the surface height better?

Data rendered with Vendor Map

Data rendered with CET-L20



Underlying data plotted as a surface

colorcet

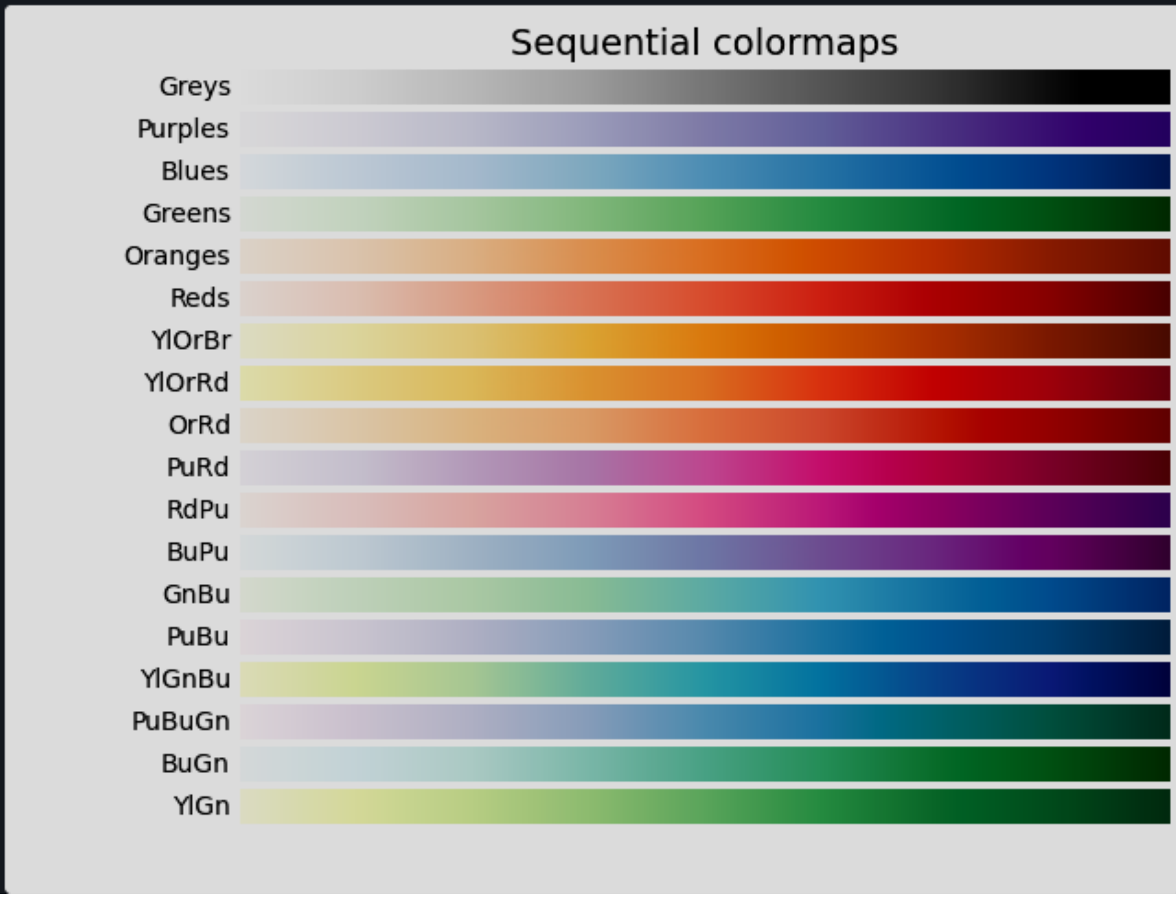
Colormap Types: Sequential/Perceptually Uniform

- [Matplotlib Website]

```
plot_color_gradients('Perceptually Uniform Sequential',
                     ['viridis', 'plasma', 'inferno', 'magma', 'cividis'])
```

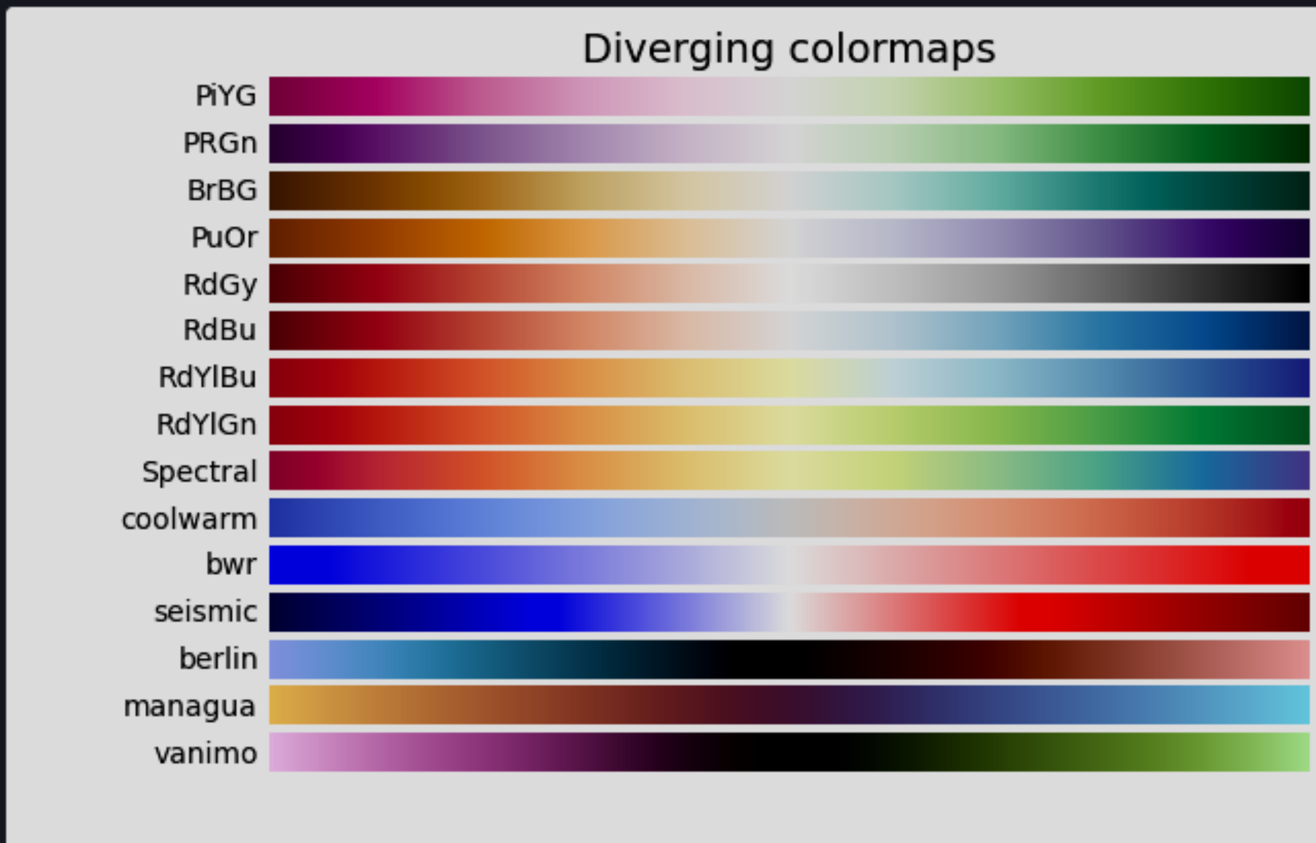
Colormap Types: Sequential

```
plot_color_gradients('Sequential',
                     ['Greys', 'Purples', 'Blues', 'Greens', 'Oranges', 'Reds',
                      'YlOrBr', 'YlOrRd', 'OrRd', 'PuRd', 'RdPu', 'BuPu',
                      'GnBu', 'PuBu', 'YlGnBu', 'PuBuGn', 'BuGn', 'YlGn'])
```



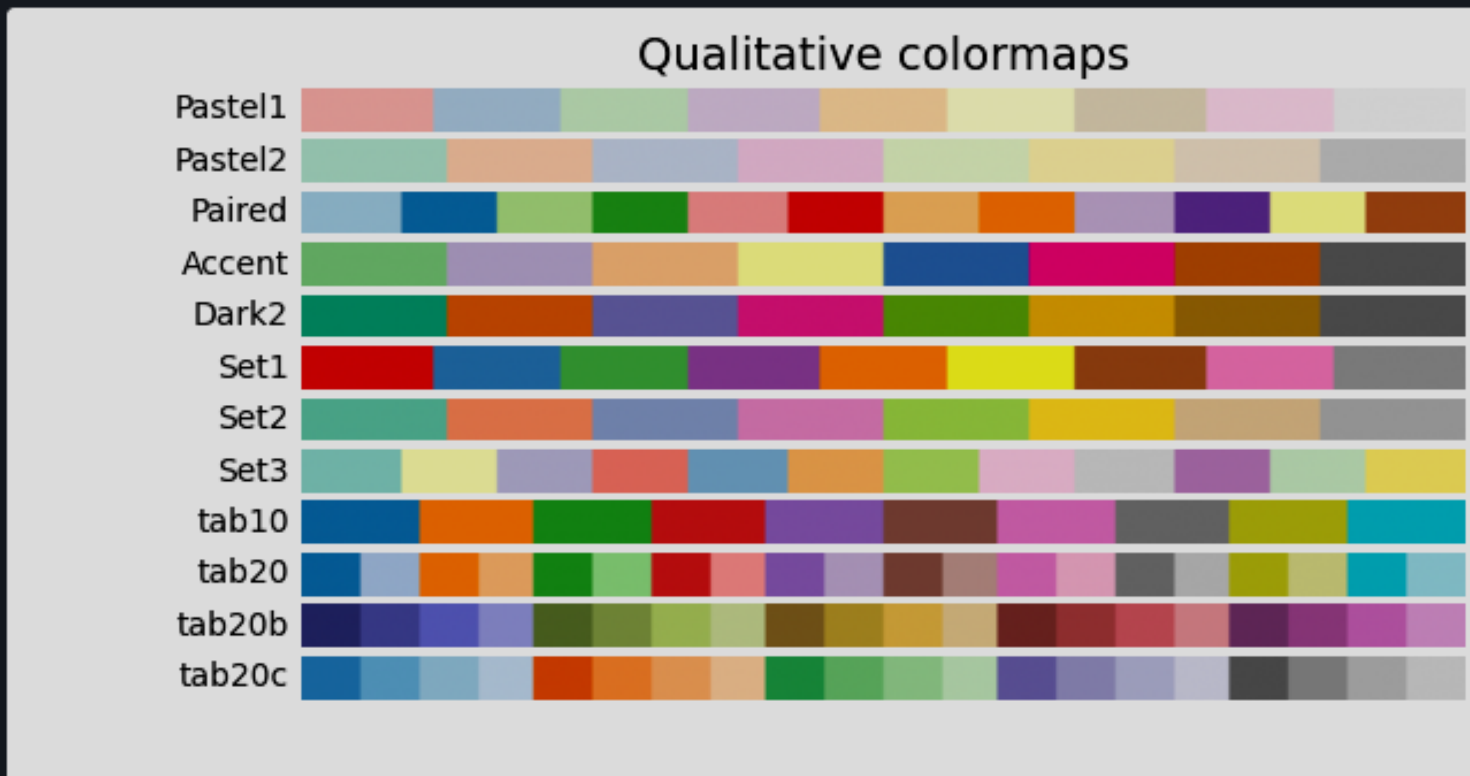
Colormap Types: Diverging

```
plot_color_gradients('Diverging',
                     ['PiYG', 'PRGn', 'BrBG', 'PuOr', 'RdGy', 'RdBu', 'RdYlBu',
                      'RdYlGn', 'Spectral', 'coolwarm', 'bwr', 'seismic',
                      'berlin', 'managua', 'vanimo'])
```



Colormap Types: Qualitative

```
plot_color_gradients('Qualitative',
                     ['Pastel1', 'Pastel2', 'Paired', 'Accent', 'Dark2',
                      'Set1', 'Set2', 'Set3', 'tab10', 'tab20', 'tab20b',
                      'tab20c'])
```



Do you need a map?

- Gelman 2004: All Maps of Parameter Estimates are Misleading

ALL MAPS OF PARAMETER ESTIMATES ARE MISLEADING

ANDREW GELMAN[1] AND PHILLIP N. PRICE[2] *

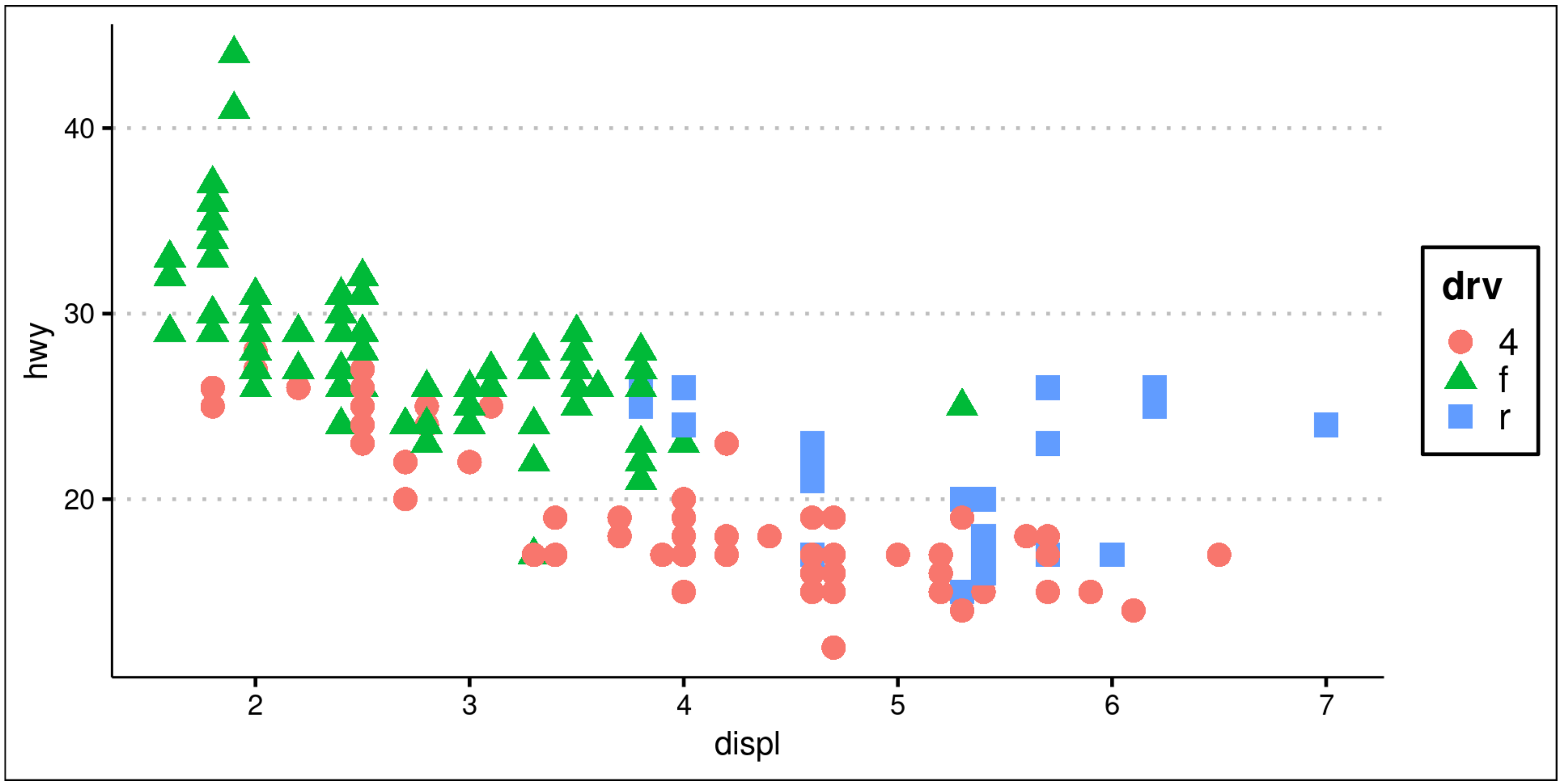
[1] *Department of Statistics, Columbia University, 618 Mathematics Building, New York, New York 10027, U.S.A.*
[2] *Lawrence Berkeley National Laboratory, LBNL 90-3058, Berkeley, California 94720, U.S.A.*

SUMMARY

Maps are frequently used to display spatial distributions of parameters of interest, such as cancer rates or average pollutant concentrations by county. It is well known that plotting observed rates can have serious drawbacks when sample sizes vary by area, since very high (and low) observed rates are found disproportionately in poorly-sampled areas. Unfortunately, adjusting the observed rates to account for the effects of small-sample noise can introduce an opposite effect, in which the highest adjusted rates tend to be found disproportionately in well-sampled areas. In either case, the maps can be difficult to interpret because the display of spatial variation in the underlying parameters of interest is confounded with spatial variation in sample sizes. As a result, spatial patterns occur in adjusted rates even if there is no spatial structure in the underlying parameters of interest, and adjusted rates tend to look too uniform in areas with little data. We introduce two models (normal and Poisson) in which parameters of interest have *no* spatial patterns, and demonstrate the existence of spatial artefacts in inference from these models. We also discuss spatial models and the extent to which they are subject to the same artefacts. We present examples from Bayesian modelling, but, as we explain, the artefacts occur generally. Copyright © 1999 John Wiley & Sons, Ltd.

Categories/Qualitative Features

- Hue/Color and Shape are good for differentiating categories
- Category limit 5 or 6



CUNY School of Professional Studies

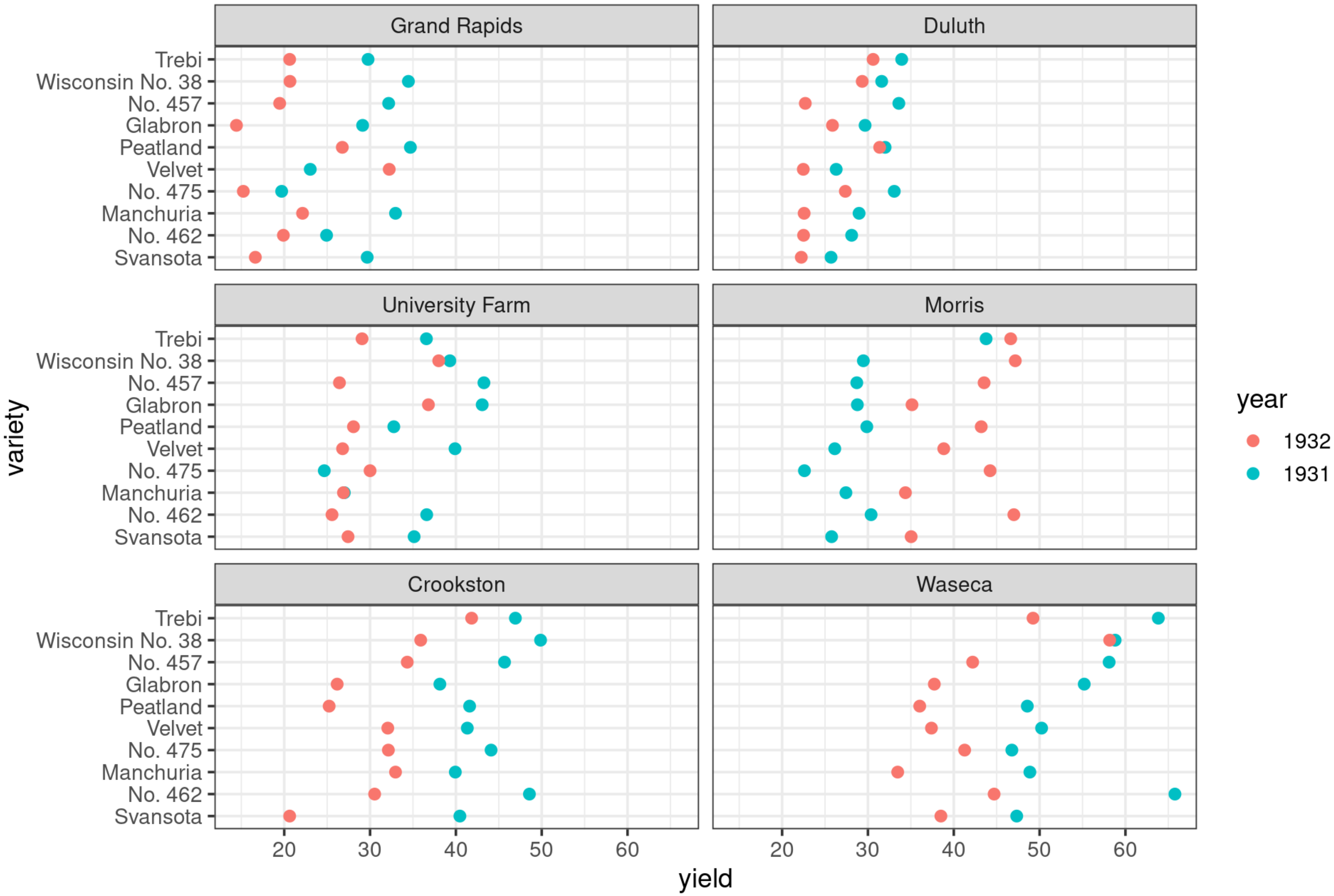
Small Multiples for Categories

- Suppose we want to show a relationship between many variables
- Make repeated plots with same axis that correspond to different categorical or numerical dimensions
- Facets, Trellis plots, small multiples plots, etc

Small Multiples for Categories

- Suppose we want to show a relationship between many variables
- Make repeated plots with same axis that correspond to different categorical or numerical dimensions
- Facets, Trellis plots, small multiples plots, etc

Small Multiples Example



CUNY School of Professional Studies

Thanks!