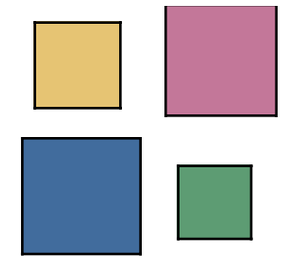


Interactive Statistical Analysis of Geographically Referenced Data

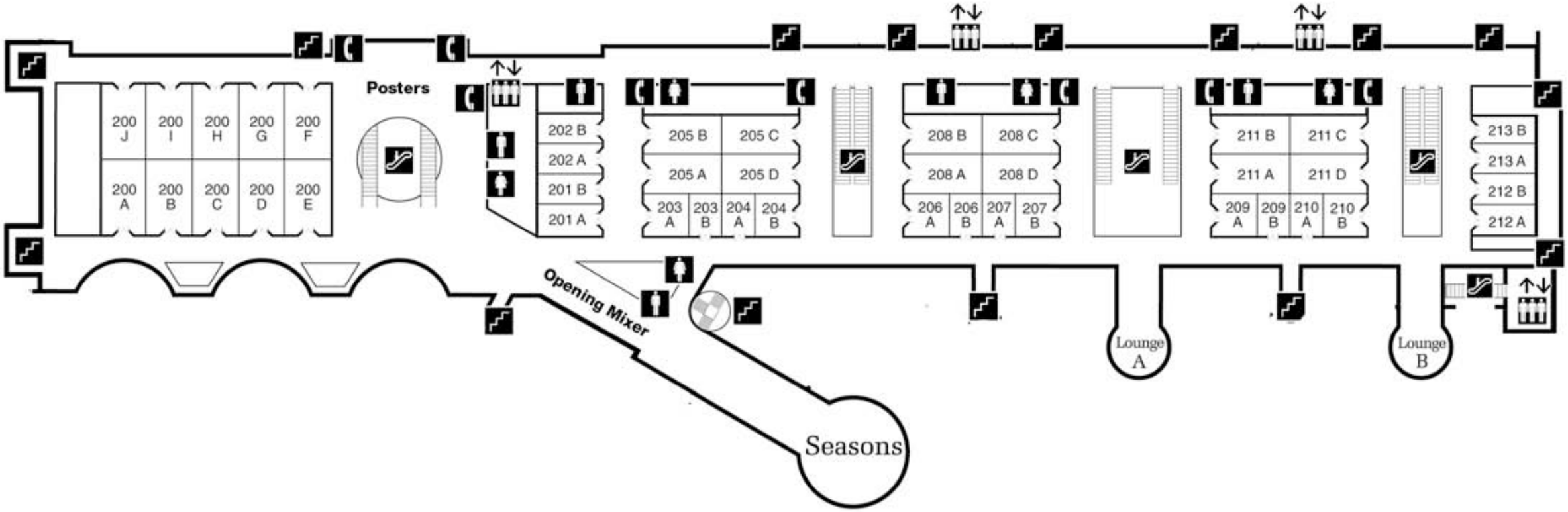
Martin Theus

Department of Computational Statistics and Data Analysis
Augsburg University

www.rosuda.org

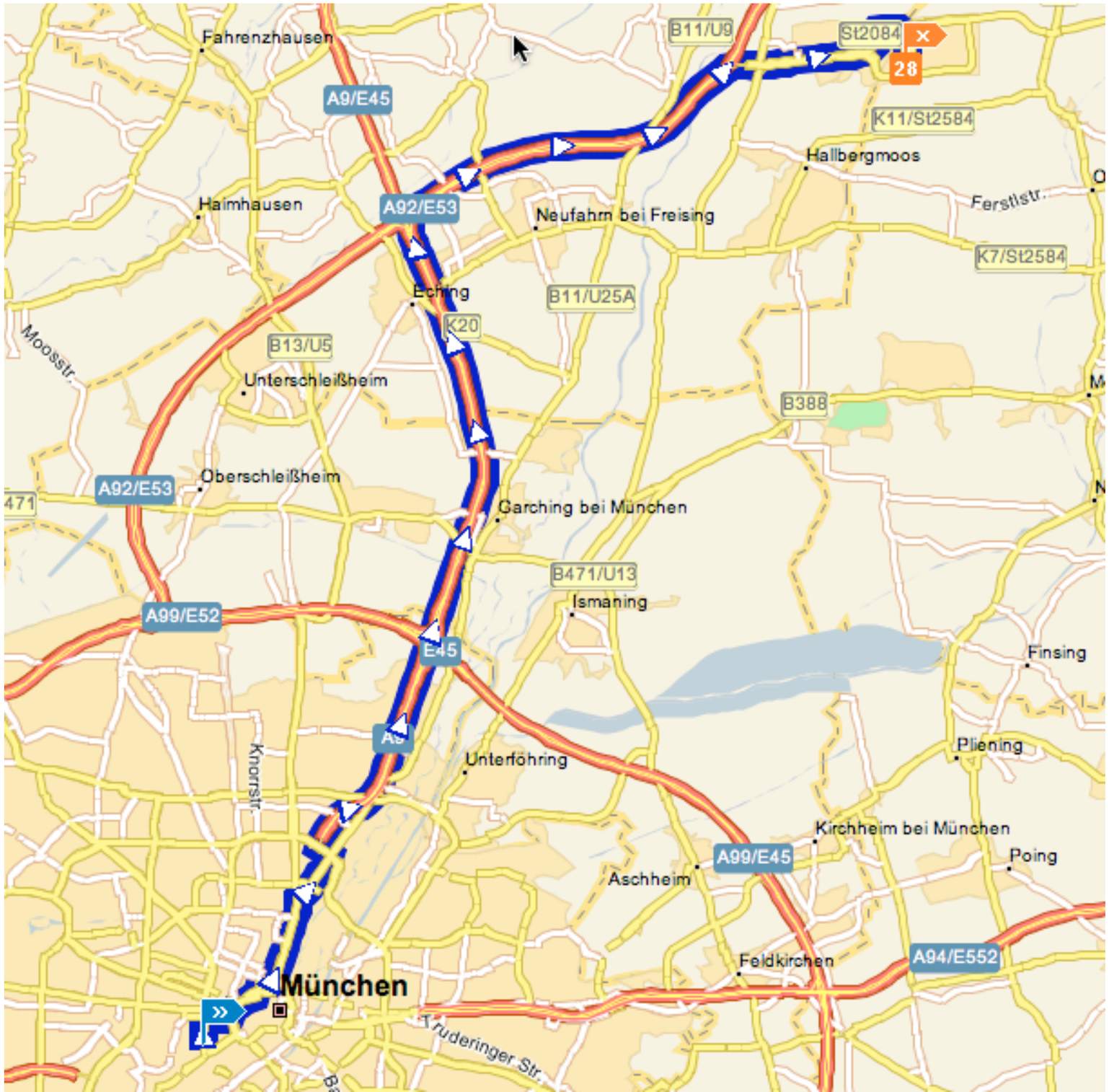
Many Kinds of Maps

JSM Floor map



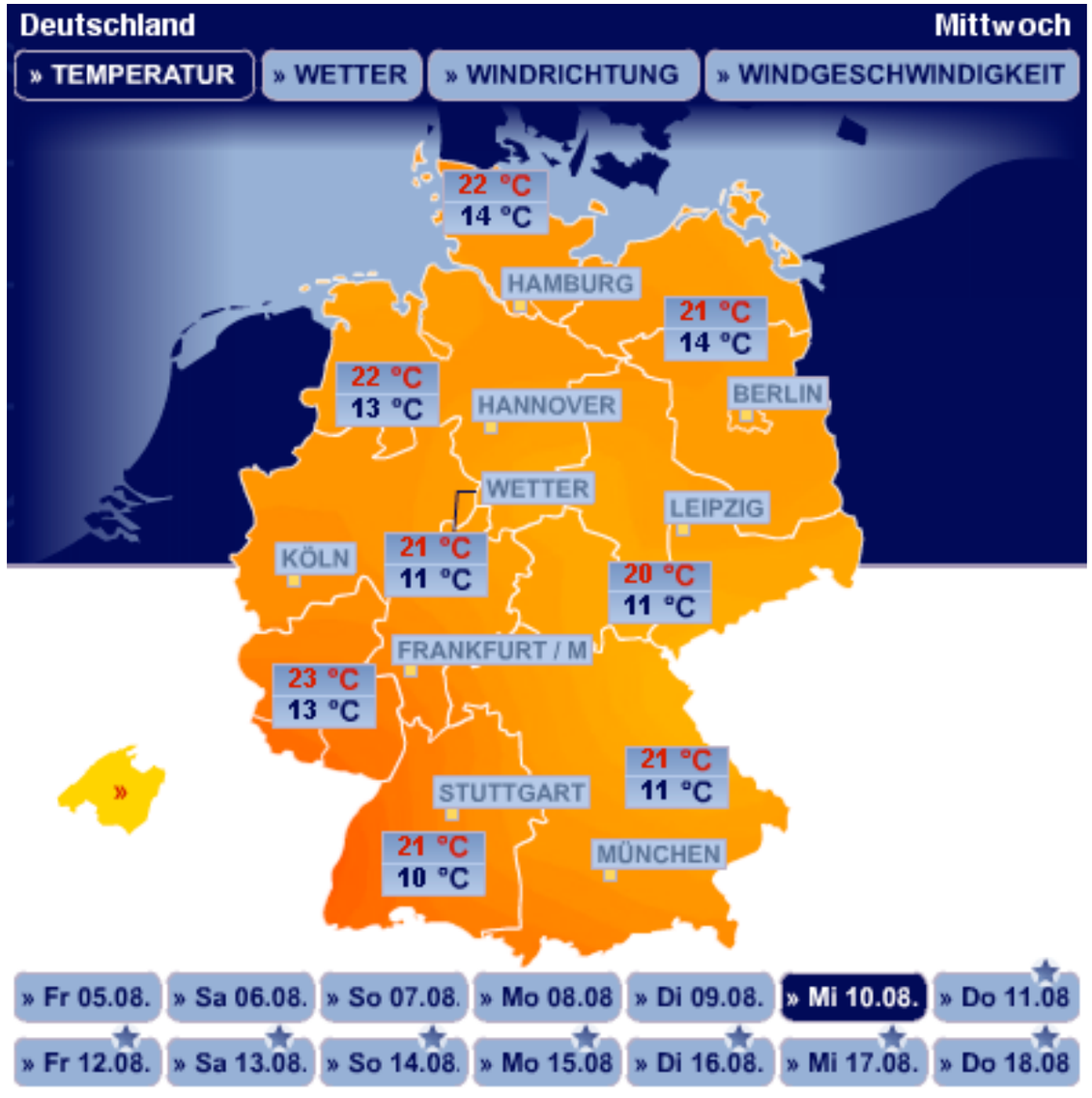
Many Kinds of Maps

Route planner



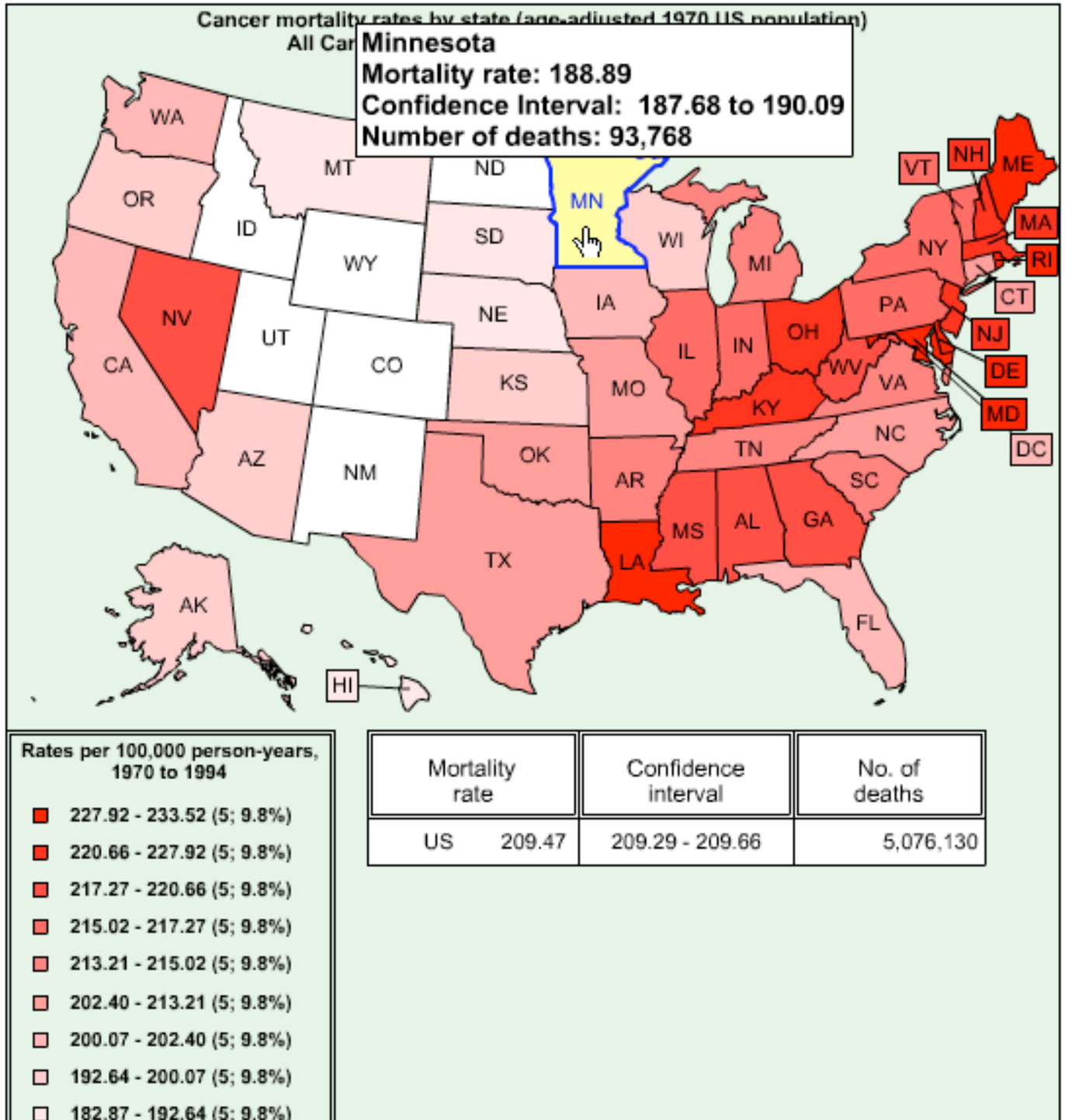
Many Kinds of Maps

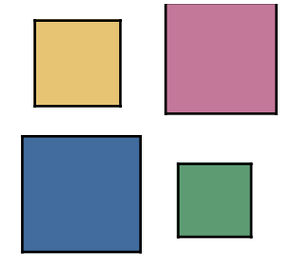
Weather maps



Many Kinds of Maps

Cancer mortality rates



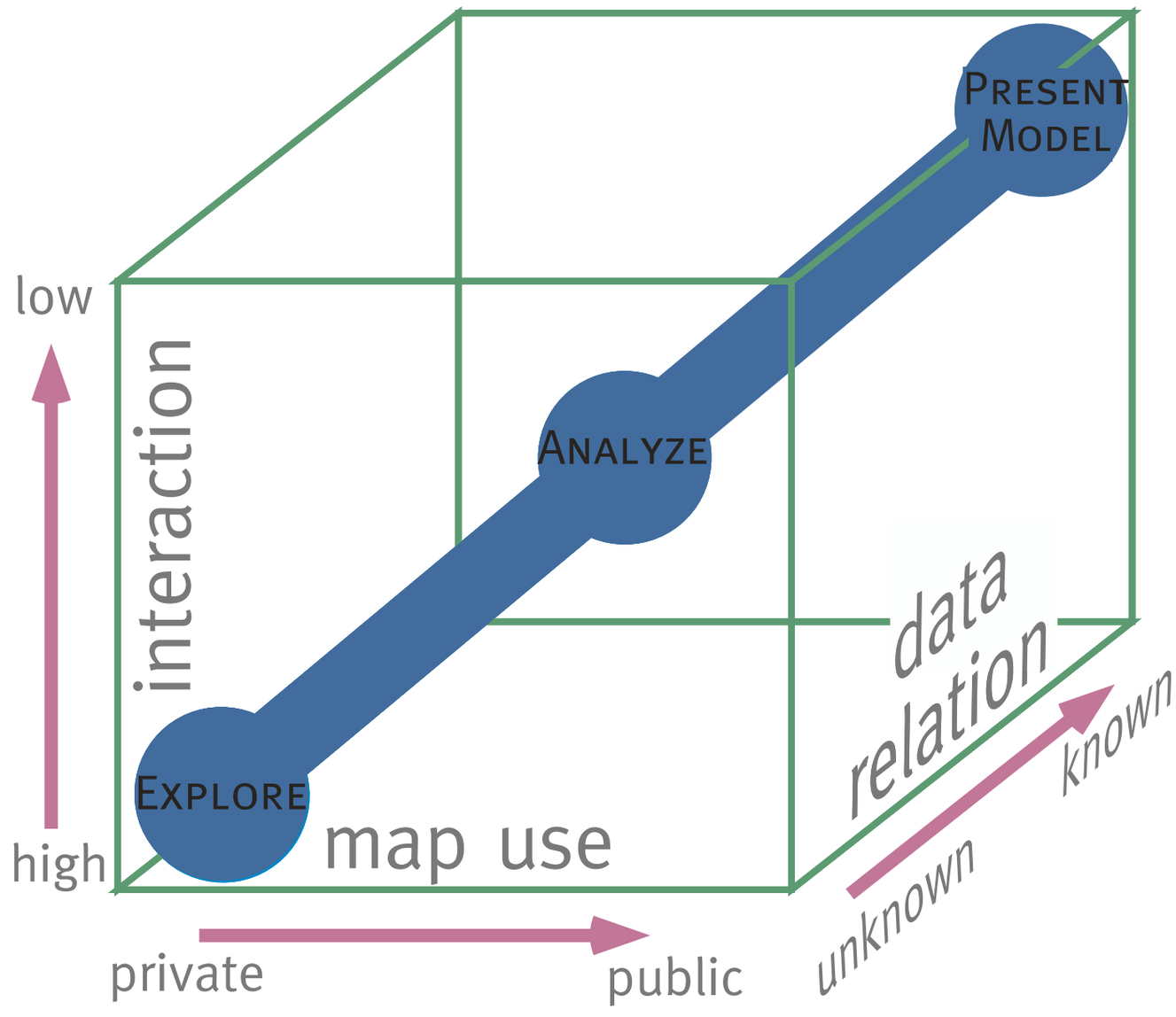


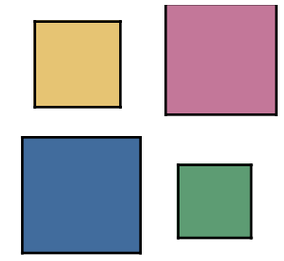
The Data vs. The Map

- Classical maps are static and were drawn by cartographers
- Typical static maps present just a single variable (at a time)
- Much care is taken to communicate this univariate information as perfectly as possible
- Today maps can be interactive, but are still “just maps”
- Adding interactivity allows the user to consider several variables simultaneously
- Ideally Exploratory Data Analysis has to be multivariate.

Presentation vs. Exploration

- Presentation
  - few variables
  - single feature focus
  - high production costs
  - maps for publication
  - little interaction
- Exploration
  - many variables
  - open ended
  - flexible
  - only for private use
  - high interaction
- Both ends come closer!



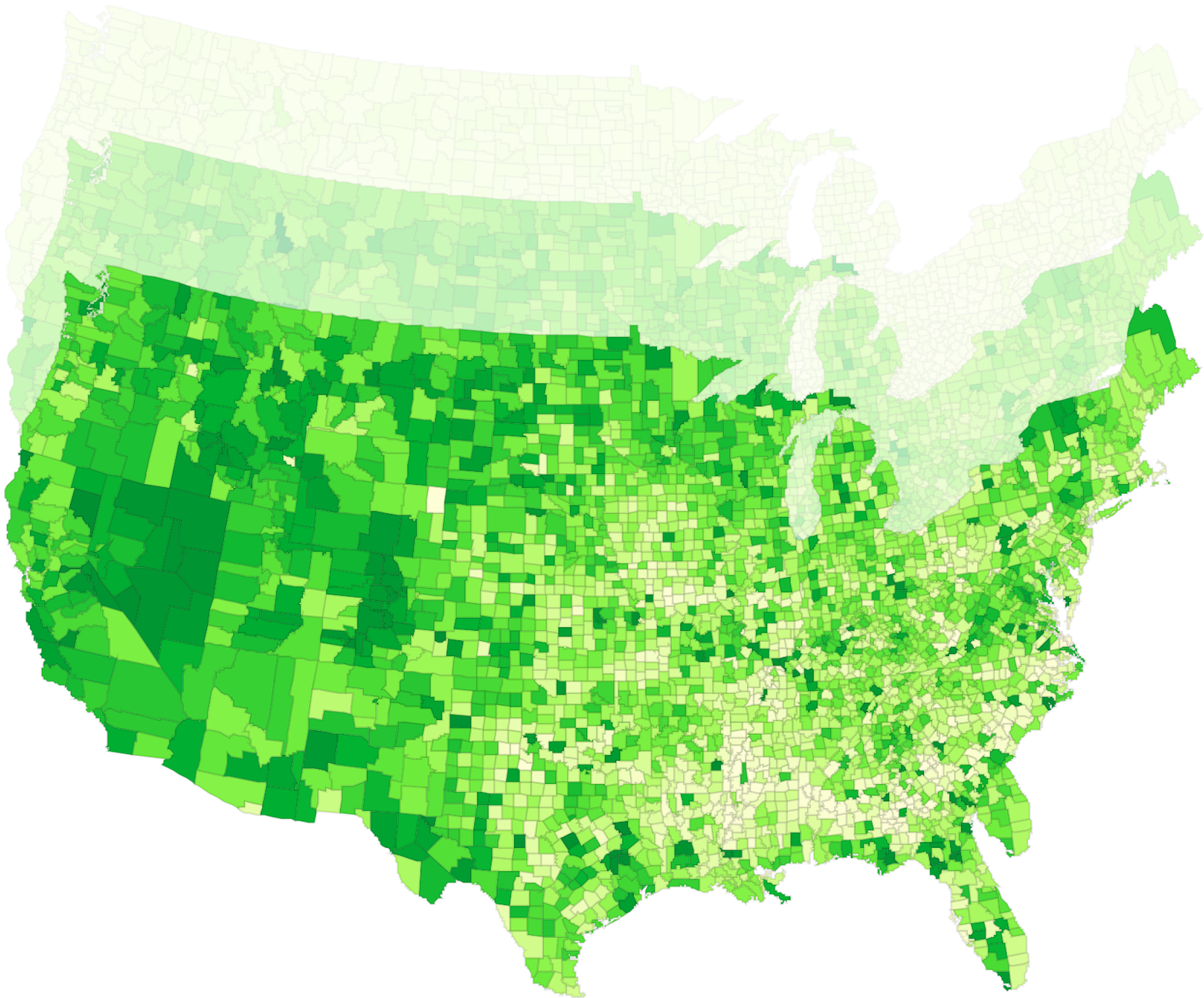


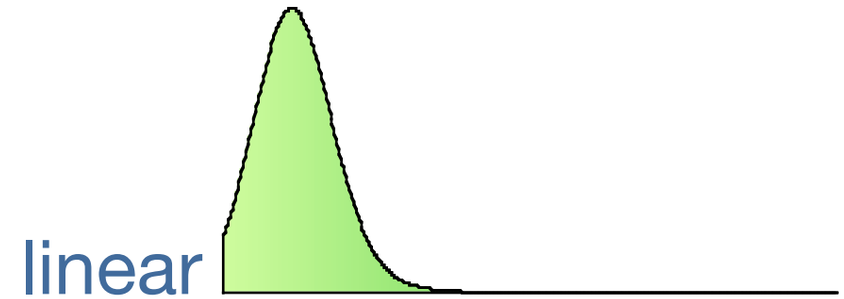
Choropleth Maps

- In choropleth maps a variable is depicted by shading each area with a corresponding color
- The mapping between values and colors is usually done linearly, the distribution of colors used in a map follows the distribution of the variable itself.
- This works well
  - for all roughly normal distributions
  - if there are no outliers present.
- For an optimal discrimination of the values depicted, the distribution of the color shadings must be close to rectangular.
- Continuous transformations can help, but transformations which are not continuously differentiable are more efficient.
  (Box-Cox is usually not sufficient)

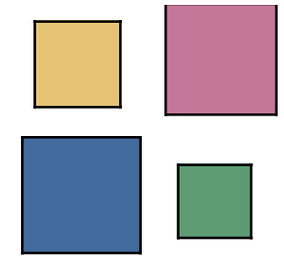
Choropleth Maps: Shading

- Question: How to shade?



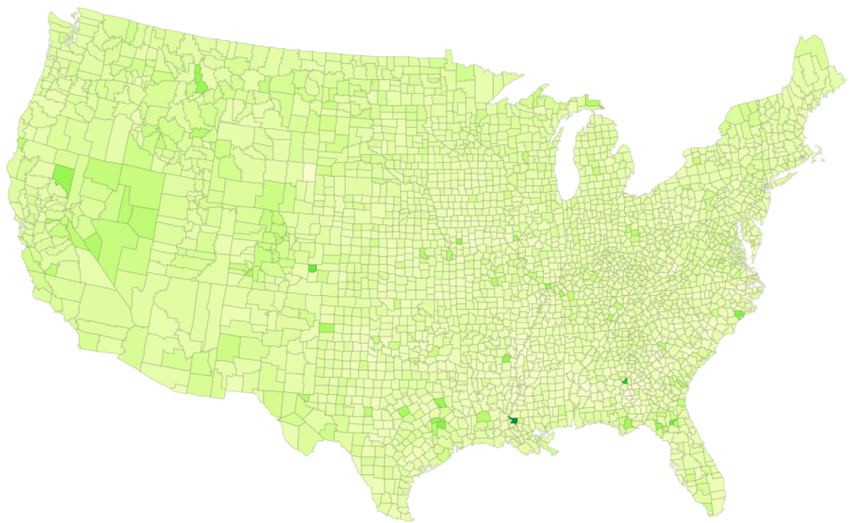


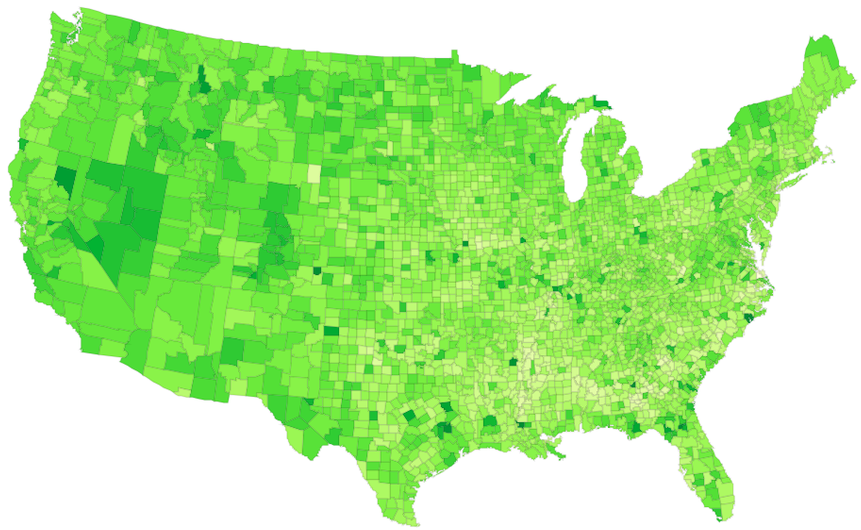
Forcing a target distribution of the colors in a choropleth map is more efficient than nonlinear transformations!

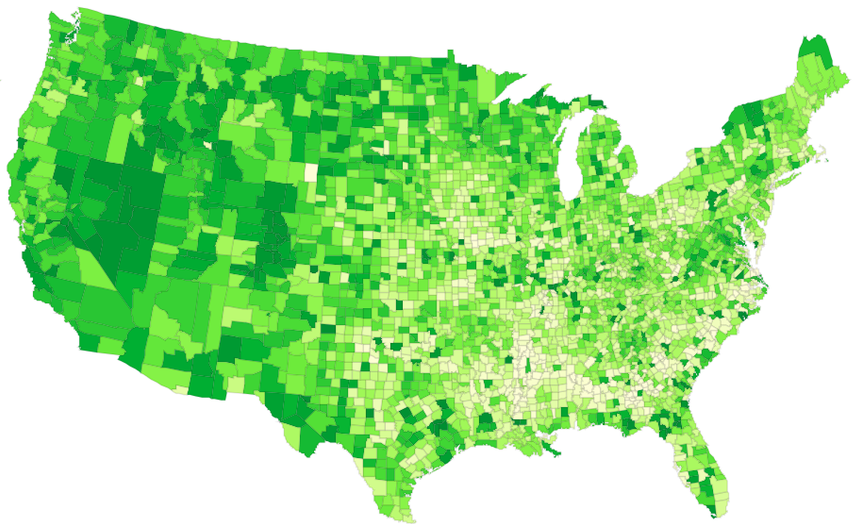


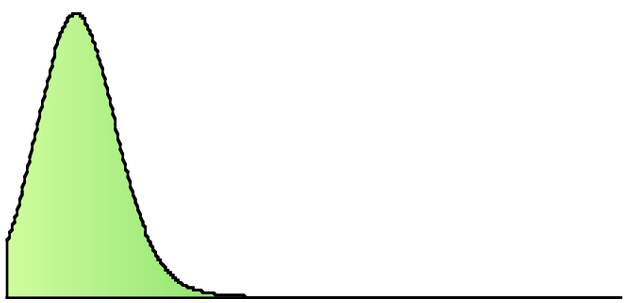
Choropleth Maps: Shading

- Comparison



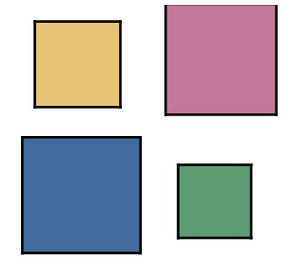




linear  normal  rank 

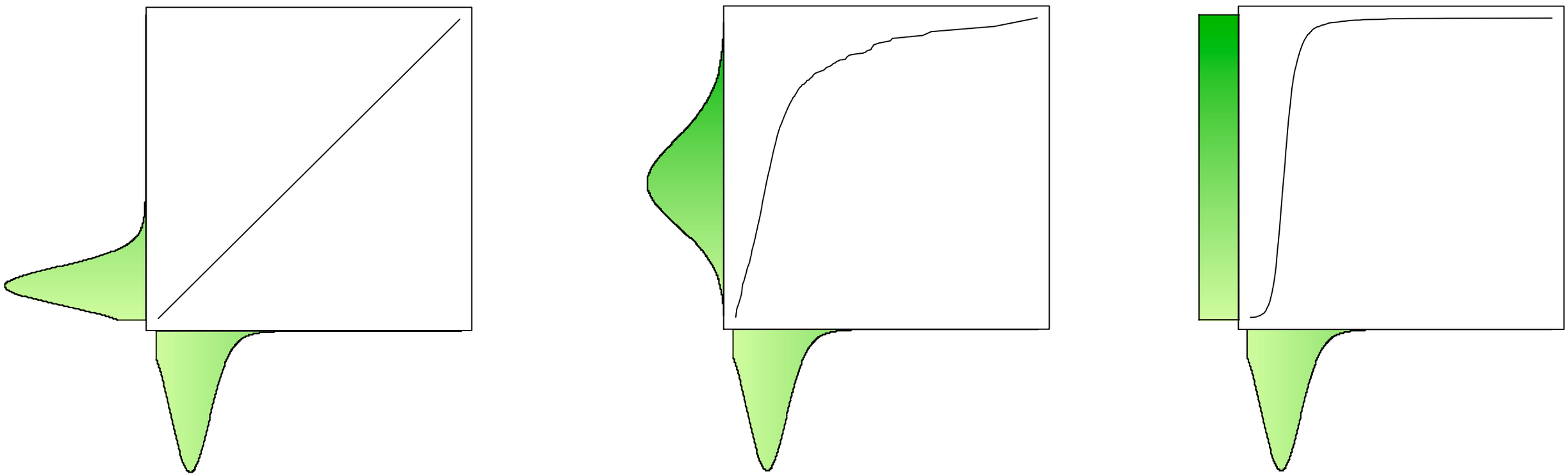
Note

- Ties must get the same color assigned
- Careful interpretation of all non-linear transformation



Choropleth Maps: Shading

- Transfer Function



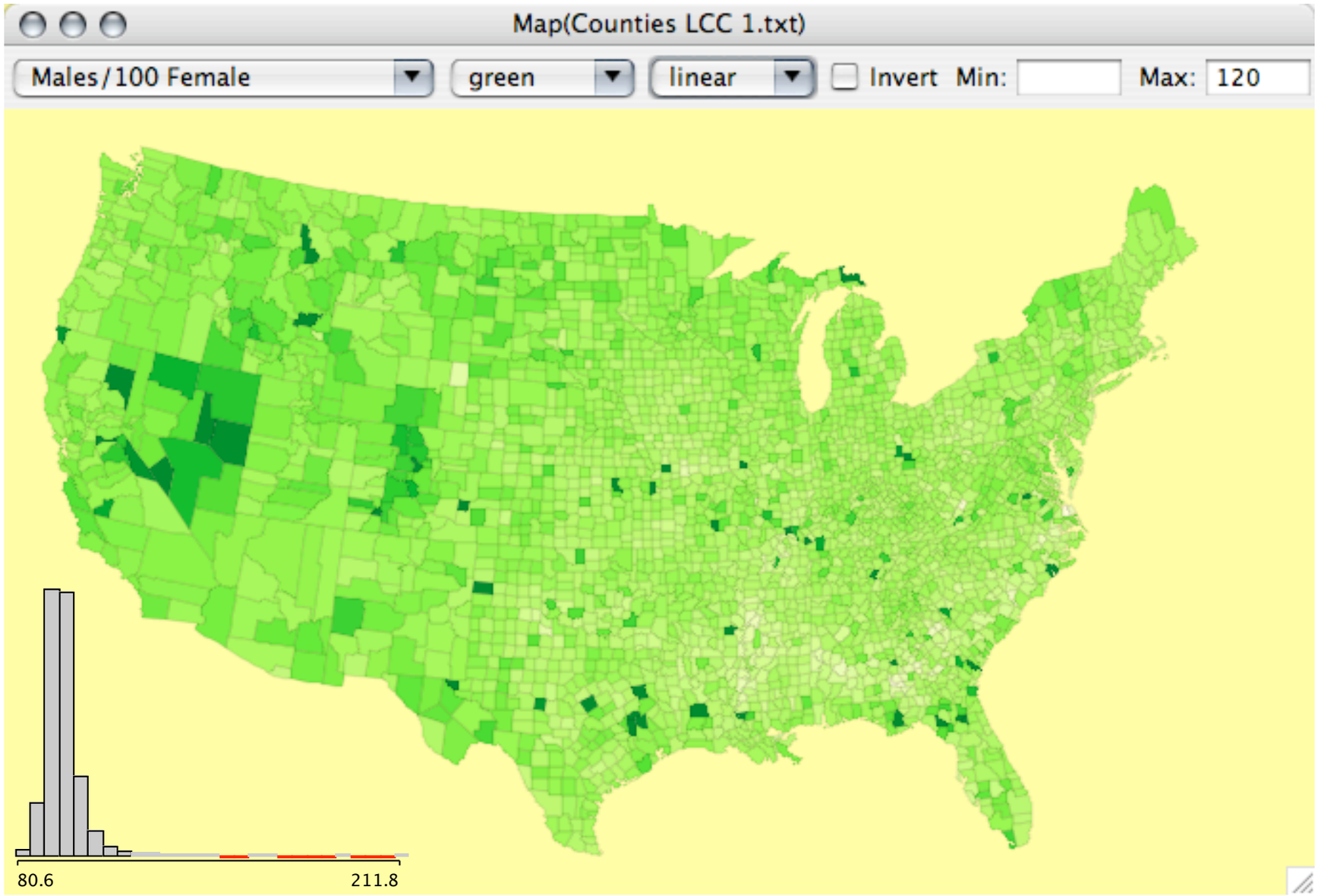
linear normal rank

Note

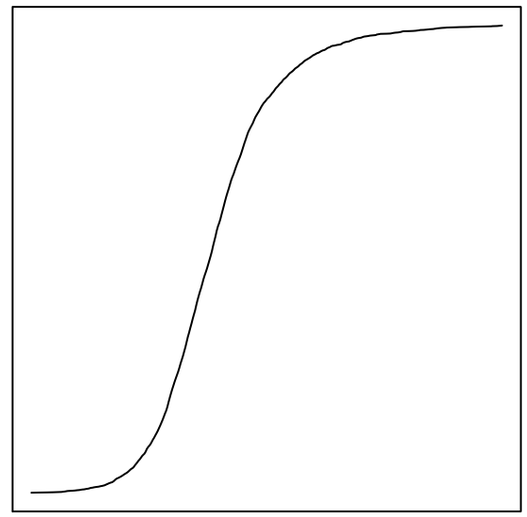
- Ties must get the same color assigned
- Careful interpretation of all non-linear transformation

Choropleth Maps: Shading (2)

- If the distribution is “only” distorted by outliers, limiting the extremes is often enough to make a linear mapping applicable.



Transfer Function for limited data

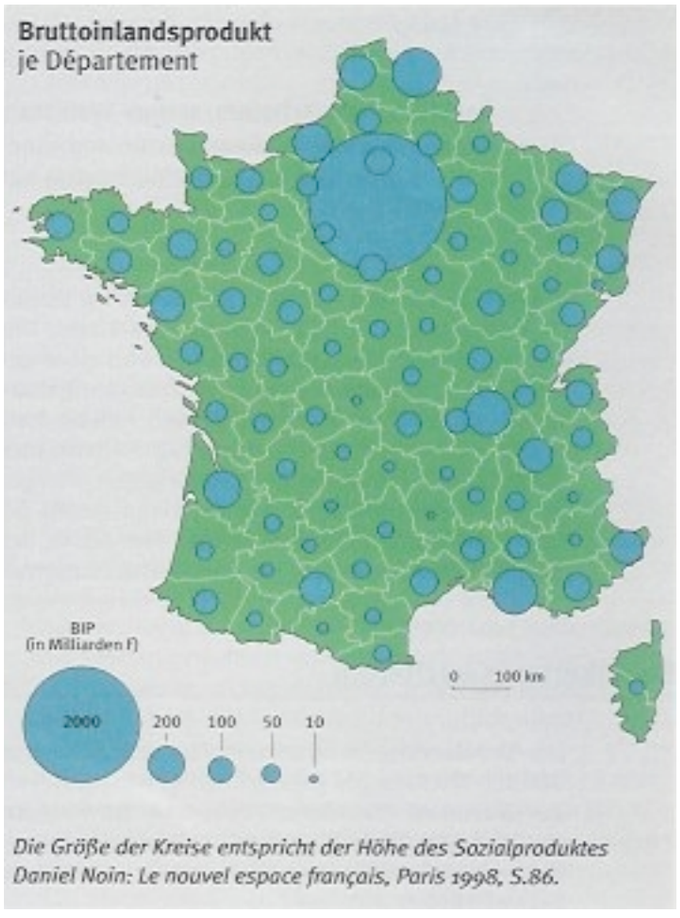


Again
Interactivity and flexibility are key!

Choropleth Maps: Alternatives

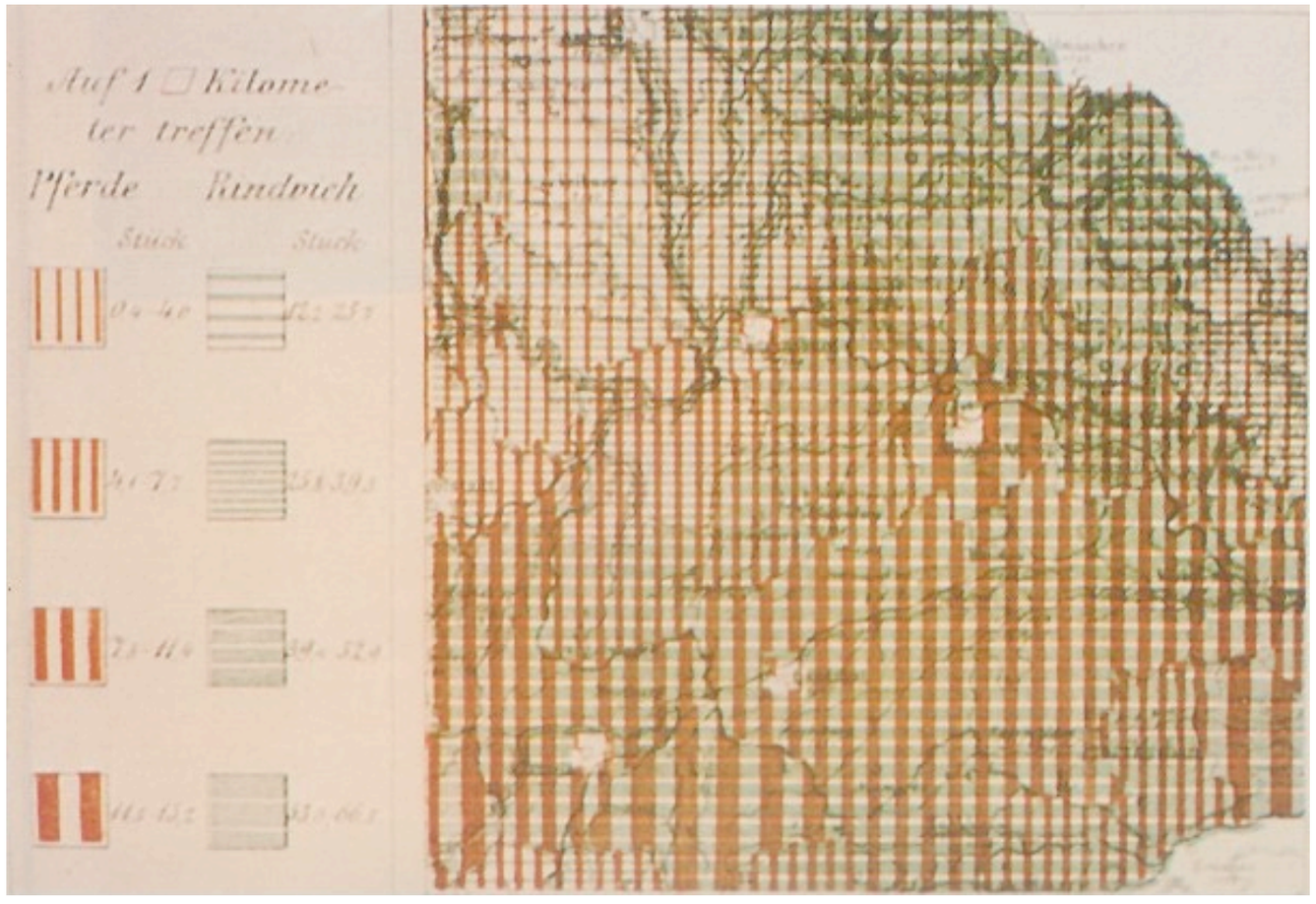
- Good alternatives are rare

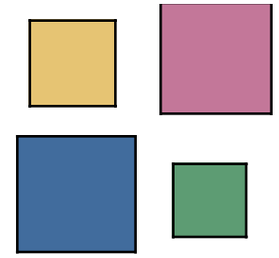
Glyphs



Die Größe der Kreise entspricht der Höhe des Sozialproduktes
Daniel Noin: Le nouvel espace français, Paris 1998, S.86.

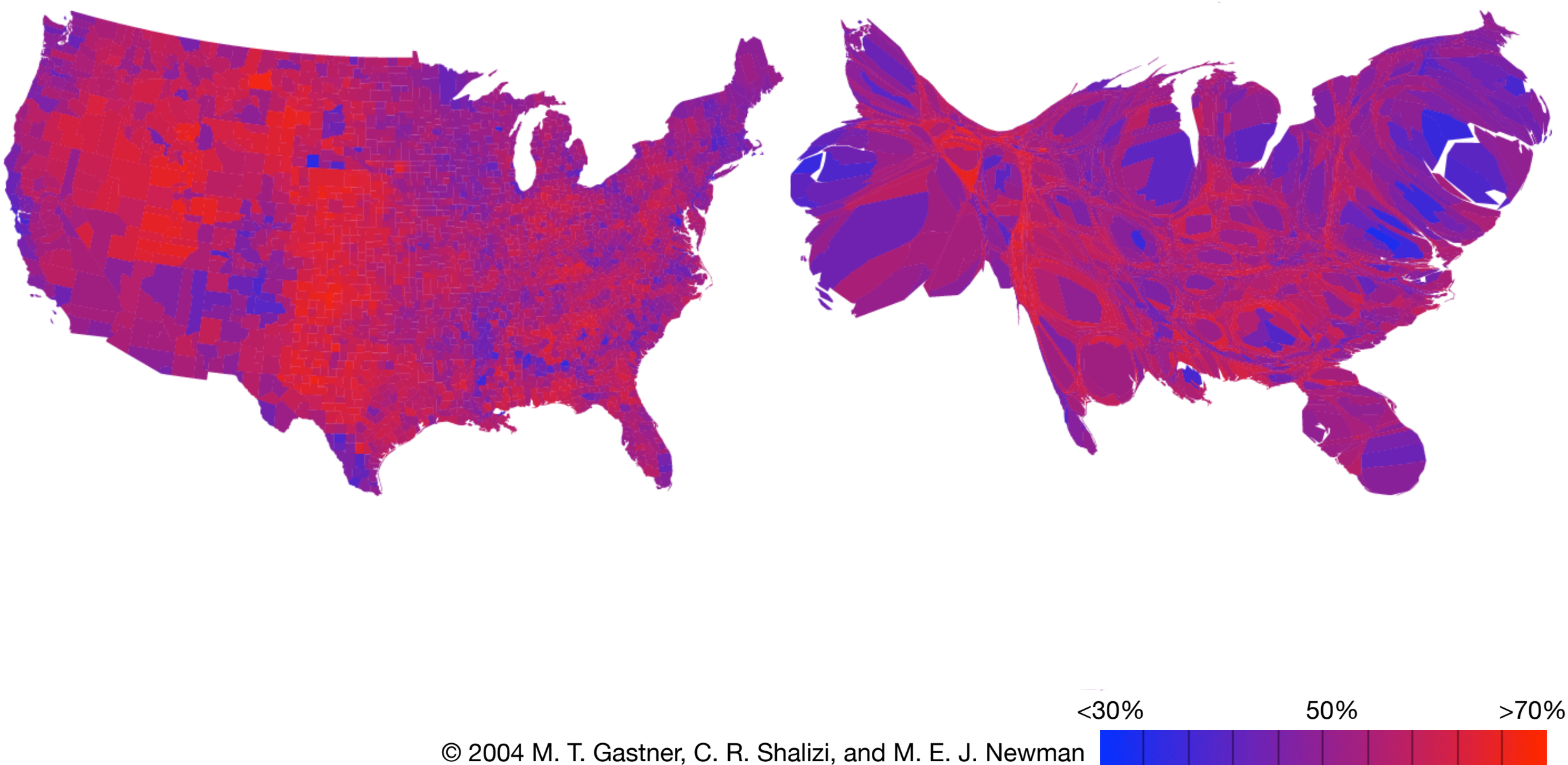
(Global) Shadings





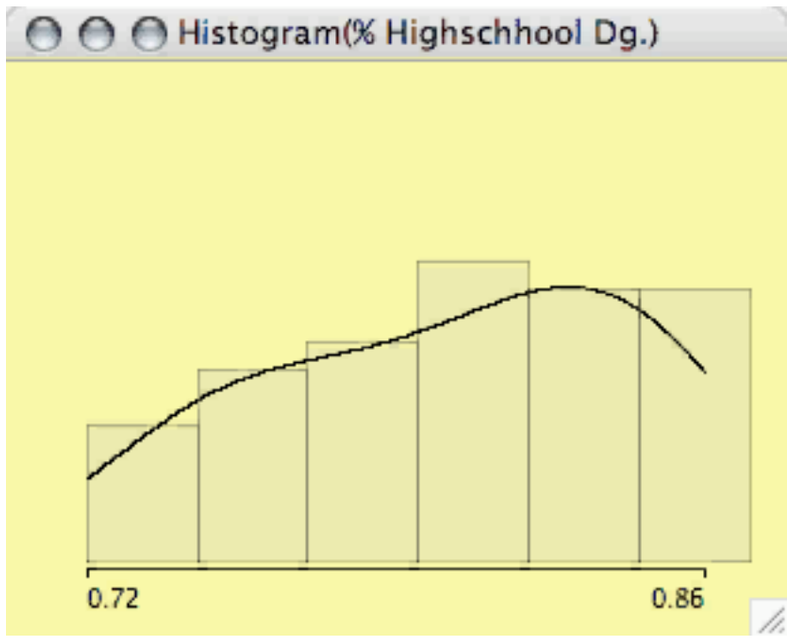
Choropleth Maps: Cartograms

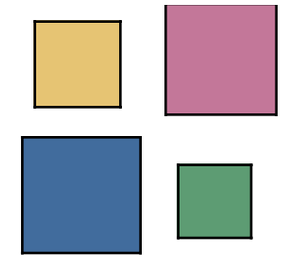
- **Problem**: Choropleth maps are biased towards large areas
  ⇒ Cartograms (2004 Election)



Choropleth Maps: Summary

- Choropleth maps seem to be the “second best” solution to display spatially referenced data, with no best solution existing
- In an interactive environment one needs to be able to look at many different views with little effort.
- Forcing the target distribution of the shading is easier than transforming the distribution of the input variable.
- Discretizing the shadings must be very well thought out – similar problems arise with histograms!

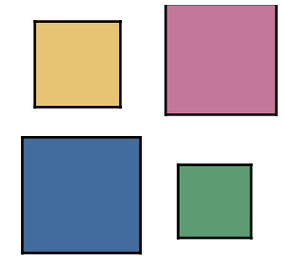




Interactions with Maps

The most important interactions with maps are:

- Resize
  Although all current desktop systems allow resizing windows, some software tools still assign fixed plotting areas.
- Zoom
  Zooming must be implemented as an interactive action. Hierarchical zooming, i.e. a zoom-back takes you to the last zoom, is desirable.
- Pan
  Moving the viewport.
- Query
  Flexible means of querying the data for a region (close to that region!)
- Shading
  - areas ✓
  - boundaries, must be adjusted to the shading of the areas!



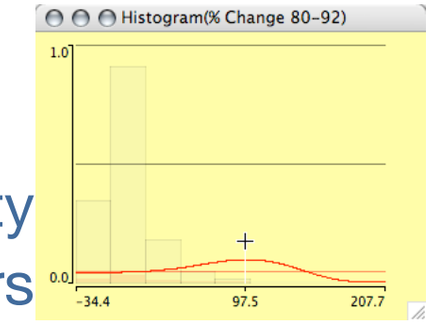
Linking the Map

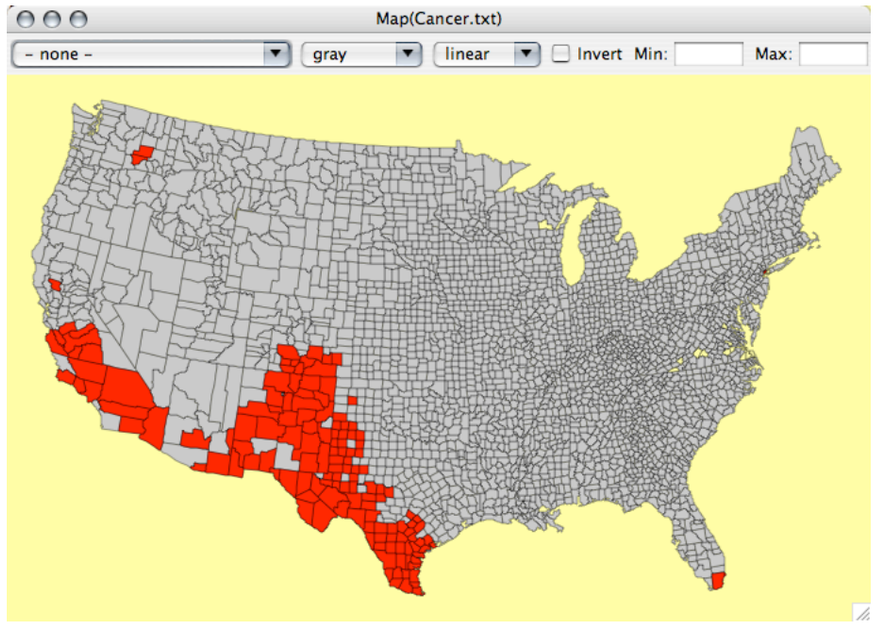
- Linking is the fundamental building block of interactive statistical graphics
- Linking a map is adding “just” another plot to the many plots we need for the statistical exploration of the data
- Objects selected in any one of the plots are highlighted in all active plots associated with the dataset
- The map serves as both lookup display and selection device
- Highlighting interferes with the shading of the map

Statistical Graphics ...

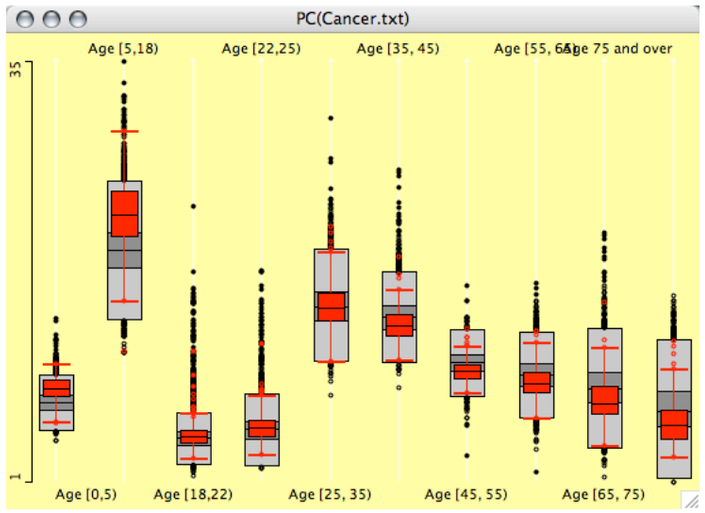
- Many statistical graphics are very useful in an interactive analysis

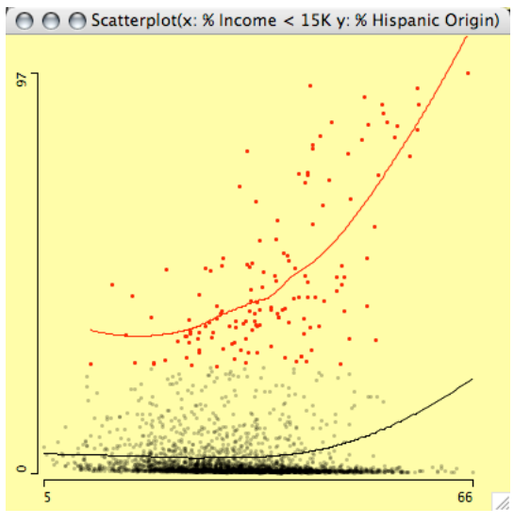
density estimators



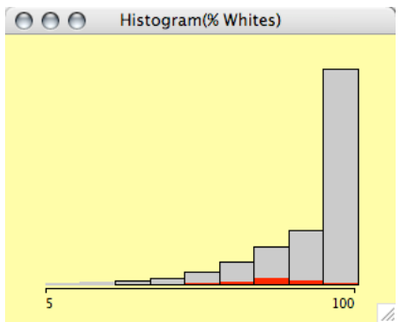


parallel box-plots

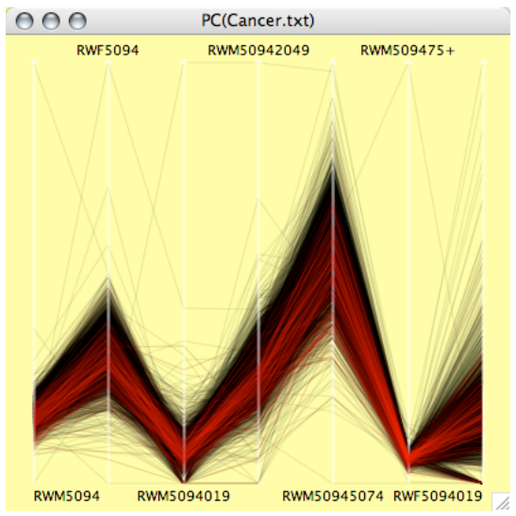




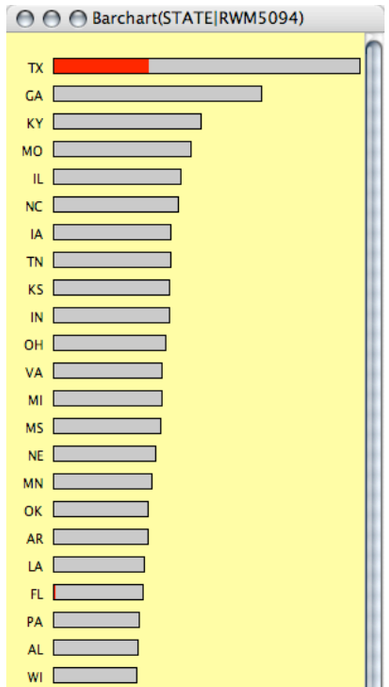
scatterplot (smoothers)



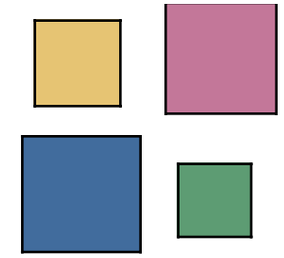
histograms



parallel coordinate plots



bar charts



Interaction with Statistical Graphics

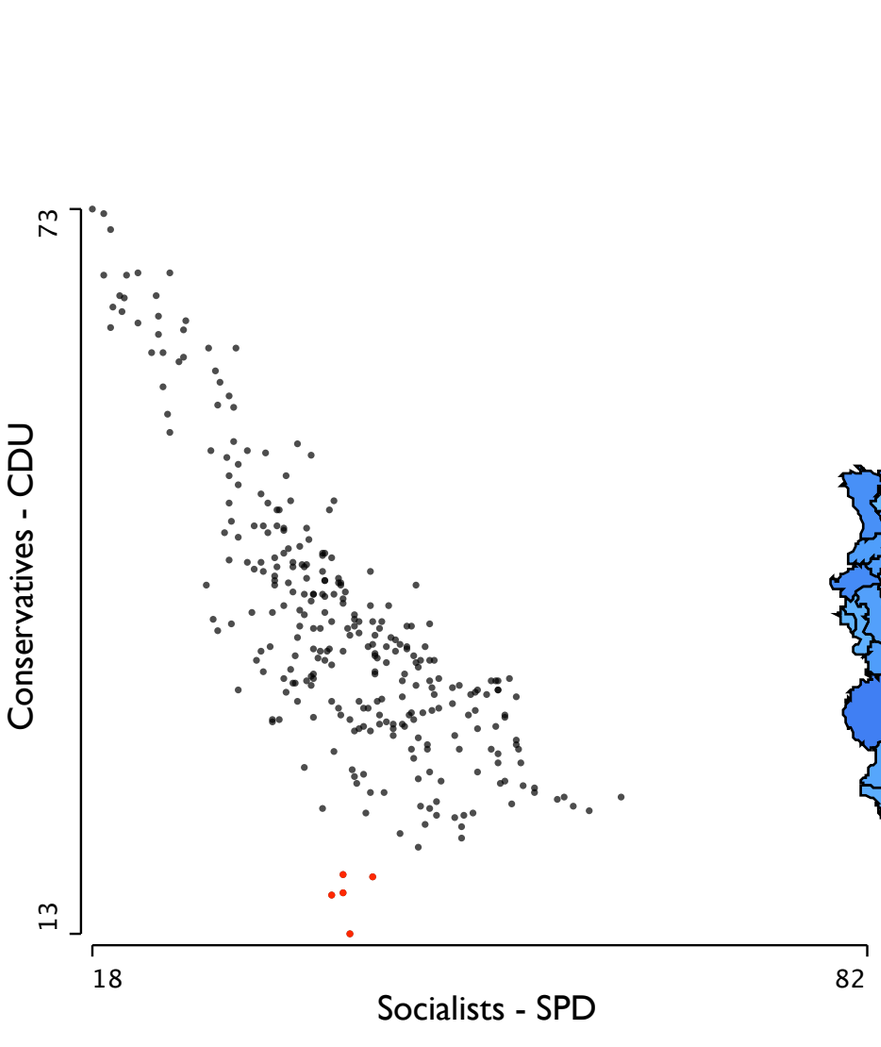
- Some interactions with maps also apply to other statistical graphics:
  - Zoom and pan (e.g. scatterplot)
  - Query, very important for **ALL** statistical graphics

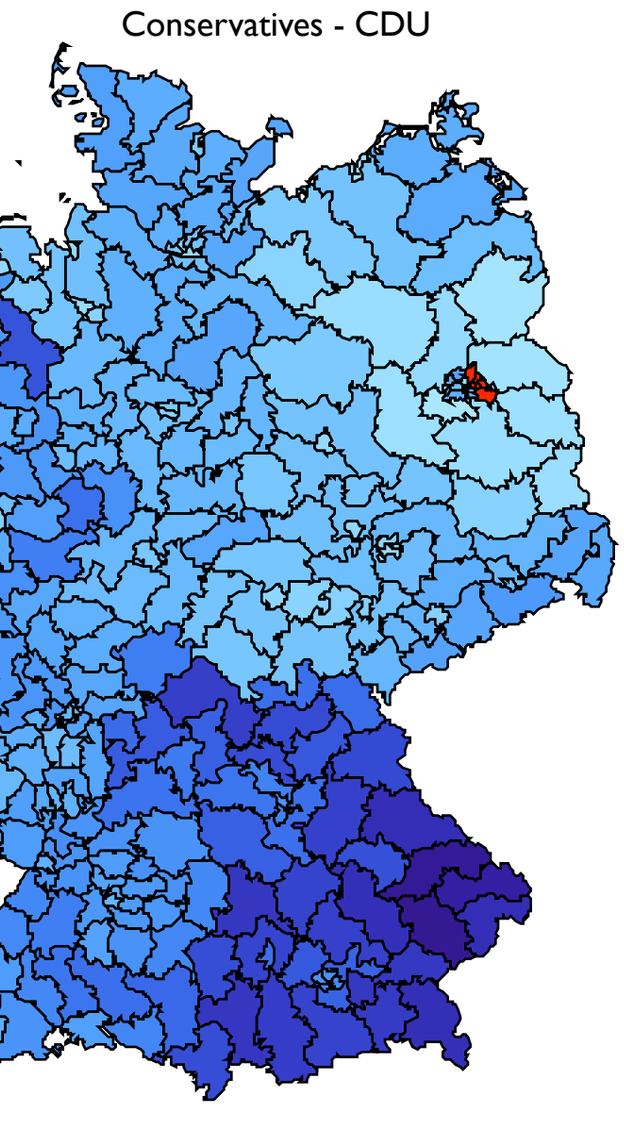
- Interactive statistical graphics offer plot specific controls like:
  - changing parameters, e.g. binwidth in histograms
  - adding / deleting axes
  - reordering of variables and categories
  - …

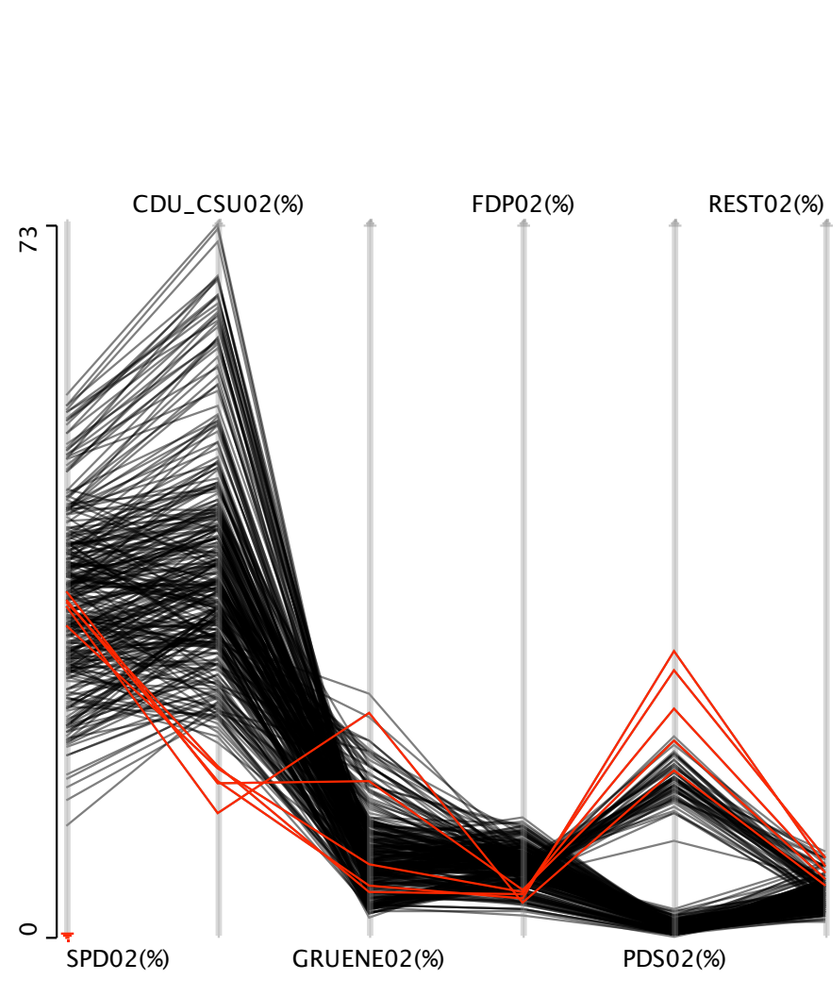
- Direct manipulation interfaces are important to support an exploratory workflow

Example: German Election Data 2002

- The former East Berlin has a distinct pattern of its own …
- The conservative candidate came from Bavaria …

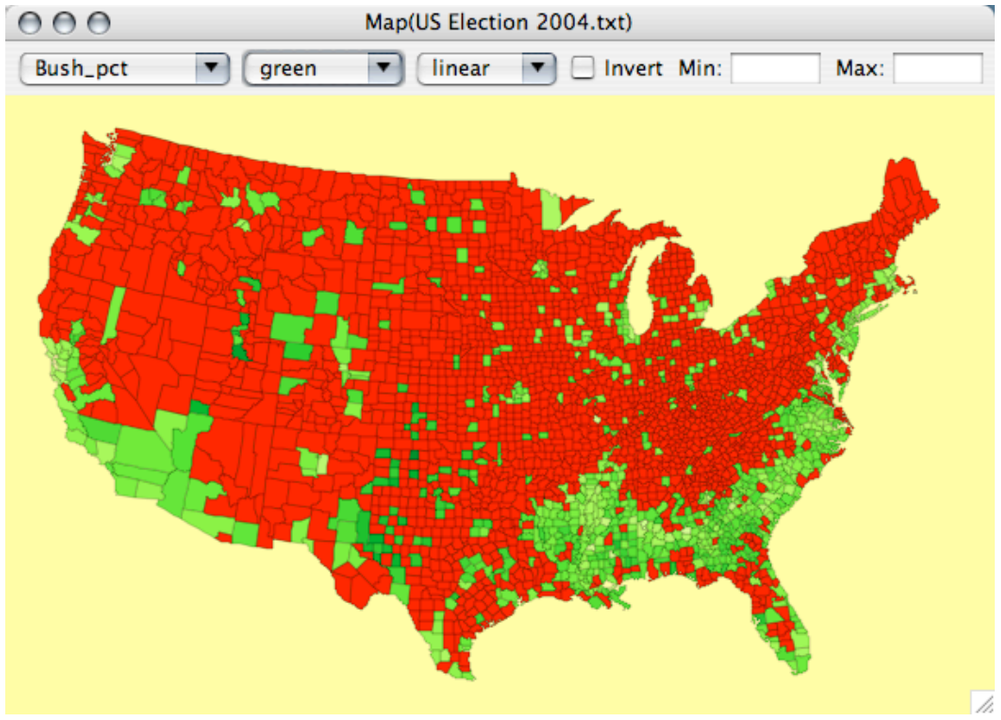


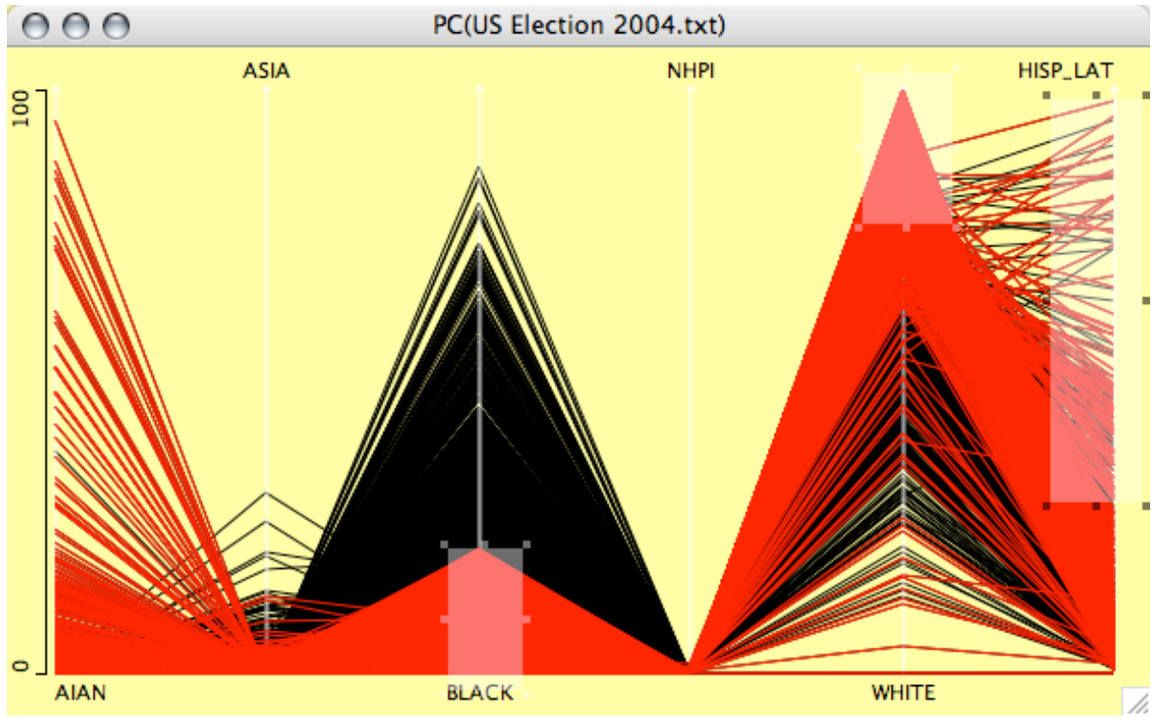


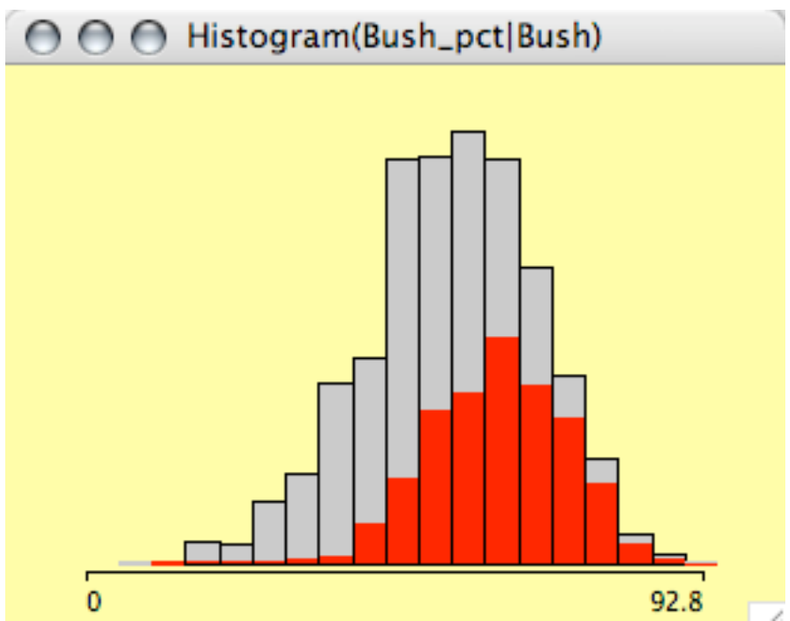


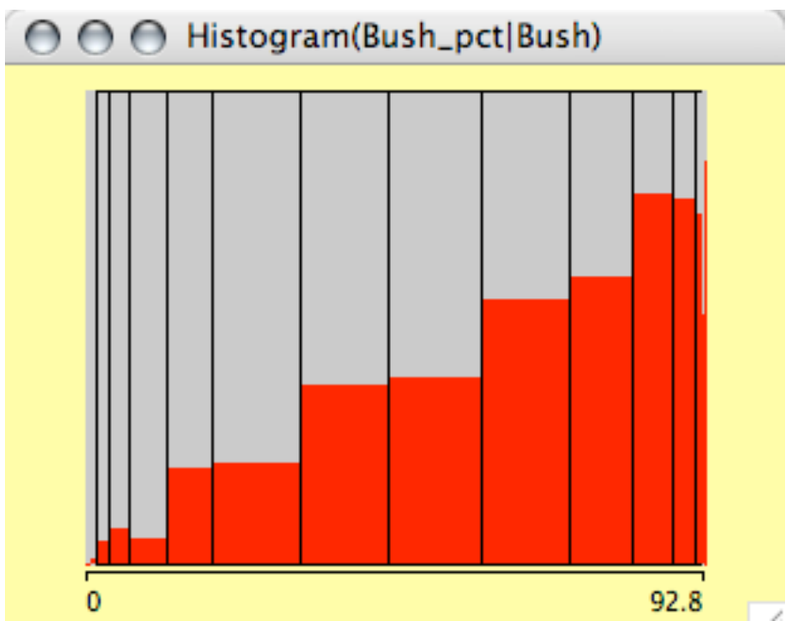
Example: US Election 2004

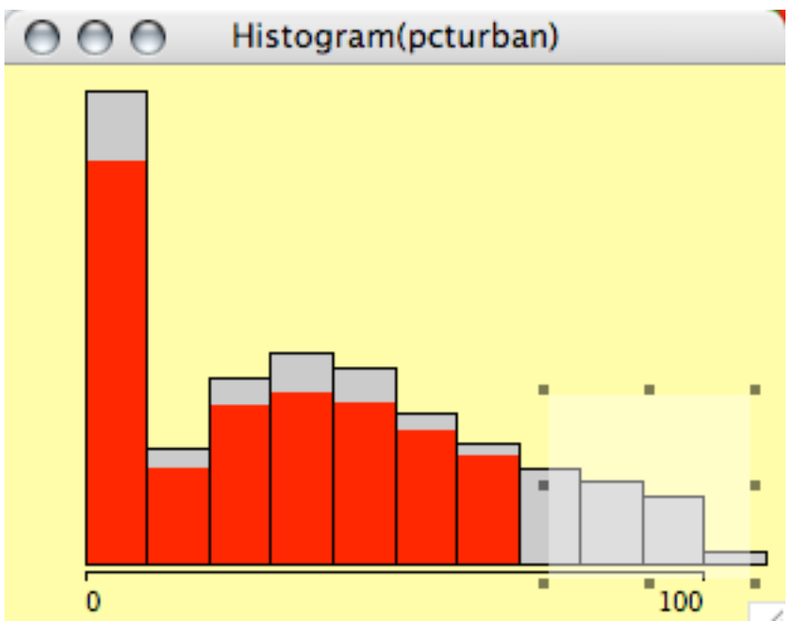
- *"Who did vote for the president?"*

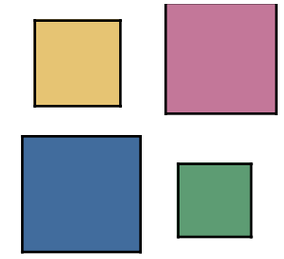












Wrap Up

- In **Exploratory Spatial Data Analysis** (ESDA), we are primarily focussed on the statistical aspects of the data, so we need (graphical) statistical tools.

- Choropleth maps for exploration differ from those used solely for presentation.

- Interactions with maps are more than just changing scale and location.

- Most of the ideas presented can be found in the **Mondrian** Software: (www.rosuda.org/Mondrian)

Thanks for your attention!