BGGN 213
Data visualization with R
Barry Grant
UC San Diego
http://thegrantlab.org/bgggn213
Recap From Last Time:

- What is R and why should we use it?
- Familiarity with R’s basic syntax.
- Familiarity with major R data structures namely vectors and data.frames.
- Understand the basics of using functions (arguments, vectorization and re-cycling).
- Be able to use R to read and parse comma-separated (.csv) formatted files ready for subsequent analysis.
- Appreciate how you can use R scripts to aid with reproducibility.

[MPA Link]
Today’s Learning Goals

• Appreciate the major elements of exploratory data analysis and why it is important to visualize data.

• Be conversant with data visualization best practices and understand how good visualizations optimize for the human visual system.

• Be able to generate informative graphical displays including scatterplots, histograms, bar graphs, boxplots, dendrograms and heatmaps and thereby gain exposure to the extensive graphical capabilities of R.

• Appreciate that you can build even more complex charts with ggplot and additional R packages such as rgl.
Today’s Learning Goals

• Appreciate the major elements of **exploratory data analysis** and why it is important to visualize data.

• Be conversant with **data visualization best practices** and understand how good visualizations optimize for the human visual system.

• Be able to generate informative graphical displays including **scatterplots**, **histograms**, **bar graphs**, **boxplots**, **dendrograms** and **heatmaps** and thereby gain exposure to the extensive graphical capabilities of R.

• Appreciate that you can build even more complex charts with **ggplot** and additional R packages such as **rgl**.
Why visualize at all?
# Over-the-Counter

## National Market System

The companies listed below reflect the volume in 100's of shares on a daily basis and the closing price and net change are reflected for the previous day's close on indices as quoted under the AMEX National Market System.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Sales Class</th>
<th>Chg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

---
<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5.00</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>4.18</td>
<td>2.75</td>
</tr>
<tr>
<td>3</td>
<td>1.98</td>
<td>4.59</td>
</tr>
<tr>
<td>4</td>
<td>-0.86</td>
<td>4.92</td>
</tr>
<tr>
<td>5</td>
<td>-3.43</td>
<td>3.64</td>
</tr>
<tr>
<td>6</td>
<td>-4.86</td>
<td>1.16</td>
</tr>
<tr>
<td>7</td>
<td>-4.70</td>
<td>-1.70</td>
</tr>
<tr>
<td>8</td>
<td>-2.99</td>
<td>-4.01</td>
</tr>
<tr>
<td>9</td>
<td>-0.30</td>
<td>-4.99</td>
</tr>
<tr>
<td>10</td>
<td>2.49</td>
<td>-4.34</td>
</tr>
<tr>
<td>11</td>
<td>4.46</td>
<td>-2.25</td>
</tr>
<tr>
<td>12</td>
<td>4.97</td>
<td>0.57</td>
</tr>
<tr>
<td>13</td>
<td>3.84</td>
<td>3.20</td>
</tr>
<tr>
<td>14</td>
<td>1.45</td>
<td>4.79</td>
</tr>
<tr>
<td>15</td>
<td>-1.42</td>
<td>4.79</td>
</tr>
<tr>
<td></td>
<td>x</td>
<td>y</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Min.</td>
<td>-4.86</td>
<td>-4.99</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-2.21</td>
<td>-1.98</td>
</tr>
<tr>
<td>Median</td>
<td>1.45</td>
<td>1.16</td>
</tr>
<tr>
<td>Mean</td>
<td>0.65</td>
<td>0.87</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>4.01</td>
<td>4.12</td>
</tr>
<tr>
<td>Max.</td>
<td>5.00</td>
<td>4.92</td>
</tr>
</tbody>
</table>
Exploratory Data Analysis

• ALWAYS look at your data!

• If you can’t see it, then don’t believe it!

• Exploratory Data Analysis (EDA) allows us to:

  1. Visualize distributions and relationships
  2. Detect errors
  3. Assess assumptions for confirmatory analysis

• EDA is the first step of data analysis!
Exploratory Data Analysis 1977

- Based on insights developed at Bell Labs in the 60’s
- Techniques for visualizing and summarizing data
- What can the data tell us? (in contrast to “confirmatory” data analysis)
- Introduced many basic techniques:
  - 5-number summary, box plots, stem and leaf diagrams,…
- 5 Number summary:
  - extremes (min and max)
  - median & quartiles
  - More robust to skewed & longtailed distributions
Chart types

- **Box-and-whisker plot**: a graphical form of 5-number summary (Tukey)
# The Trouble with Summary Stats

<table>
<thead>
<tr>
<th>Set A</th>
<th>Set B</th>
<th>Set C</th>
<th>Set D</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>8.04</td>
<td>10</td>
<td>9.14</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8</td>
<td>8.14</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>13</td>
<td>8.74</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>9</td>
<td>8.77</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>11</td>
<td>9.26</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>14</td>
<td>8.1</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6</td>
<td>6.13</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>4</td>
<td>3.1</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>12</td>
<td>9.11</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7</td>
<td>7.26</td>
</tr>
<tr>
<td>5</td>
<td>5.68</td>
<td>5</td>
<td>4.74</td>
</tr>
</tbody>
</table>

**Summary Statistics Linear Regression**

\[
u_X = 9.0 \quad \sigma_X = 3.317 \quad Y = 3 + 0.5X
\]

\[
u_Y = 7.5 \quad \sigma_Y = 2.03 \quad R^2 = 0.67
\]

[Anscombe 73]
Key point: You need to visualize your data!
Today’s Learning Goals

- Appreciate the major elements of exploratory data analysis and why it is important to visualize data.

- Be conversant with data visualization best practices and understand how good visualizations optimize for the human visual system.

- Be able to generate informative graphical displays including scatterplots, histograms, bar graphs, boxplots, dendrograms and heatmaps and thereby gain exposure to the extensive graphical capabilities of R.

- Appreciate that you can build even more complex charts with ggplot and additional R packages such as rgl.
The Elements of Graphing Data

William S. Cleveland
The Visual Display of Quantitative Information

EDWARD R. TUFTE
Key Point:
Good visualizations optimize for the human visual system.
Key Point: The most important measurement should exploit the highest ranked encoding possible

- Position along a common scale
- Position on identical but nonaligned scales
- Length
- Angle or Slope
- Area
- Volume or Density or Color saturation/hue
Key Point: The most important measurement should exploit the highest ranked encoding possible

- Position along a common scale
- Position on identical but nonaligned scales
- Length
- Angle or Slope
- Area
- Volume or Density or Color saturation/hue
Key Point: The most important measurement should exploit the highest ranked encoding possible

- Position along a common scale
- Position on identical but nonaligned scales
- Length
- Angle or Slope
- Area
- Volume or Density or Color saturation/hue
Observation: Alphabetical is almost never the correct ordering of a categorical variable.
<table>
<thead>
<tr>
<th>Car Model</th>
<th>MPG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toyota Corolla</td>
<td>30</td>
</tr>
<tr>
<td>Fiat 128</td>
<td>25</td>
</tr>
<tr>
<td>Lotus Europa</td>
<td>20</td>
</tr>
<tr>
<td>Honda Civic</td>
<td>15</td>
</tr>
<tr>
<td>Fiat X1-9</td>
<td>10</td>
</tr>
<tr>
<td>Porsche 914-2</td>
<td></td>
</tr>
<tr>
<td>Merc 240D</td>
<td></td>
</tr>
<tr>
<td>Merc 230</td>
<td></td>
</tr>
<tr>
<td>Datsun 710</td>
<td></td>
</tr>
<tr>
<td>Toyota Corona</td>
<td></td>
</tr>
<tr>
<td>Volvo 142E</td>
<td></td>
</tr>
<tr>
<td>Hornet 4 Drive</td>
<td></td>
</tr>
<tr>
<td>Mazda RX4 Wag</td>
<td></td>
</tr>
<tr>
<td>Mazda RX4</td>
<td></td>
</tr>
<tr>
<td>Ferrari Dino</td>
<td></td>
</tr>
<tr>
<td>Pontiac Firebird</td>
<td></td>
</tr>
<tr>
<td>Merc 280</td>
<td></td>
</tr>
<tr>
<td>Hornet Sportabout</td>
<td></td>
</tr>
<tr>
<td>Valiant</td>
<td></td>
</tr>
<tr>
<td>Merc 280C</td>
<td></td>
</tr>
<tr>
<td>Merc 450SL</td>
<td></td>
</tr>
<tr>
<td>Merc 450SE</td>
<td></td>
</tr>
<tr>
<td>Ford Pantera L</td>
<td></td>
</tr>
<tr>
<td>Dodge Challenger</td>
<td></td>
</tr>
<tr>
<td>Merc 450SLC</td>
<td></td>
</tr>
<tr>
<td>AMC Javelin</td>
<td></td>
</tr>
<tr>
<td>Maserati Bora</td>
<td></td>
</tr>
<tr>
<td>Chrysler Imperial</td>
<td></td>
</tr>
<tr>
<td>Duster 360</td>
<td></td>
</tr>
<tr>
<td>Camaro Z28</td>
<td></td>
</tr>
<tr>
<td>Lincoln Continental</td>
<td></td>
</tr>
<tr>
<td>Cadillac Fleetwood</td>
<td></td>
</tr>
</tbody>
</table>

mpg
The most important measurement should exploit the highest ranked encoding possible.

- Position along a common scale
- Position on identical but nonaligned scales
- Length
- Angle or Slope
- **Area**
- Volume or Density or Color saturation/hue
The most important measurement should exploit the highest ranked encoding possible.

- Position along a common scale
- Position on identical but nonaligned scales
- Length
- Angle or Slope
- Area
- Volume or Density or Color saturation/hue
<table>
<thead>
<tr>
<th>Cadillac Fleet</th>
<th>Lincoln Continental</th>
<th>Camaro Z28</th>
<th>Duster 360</th>
<th>Chrysler Imp</th>
<th>Maserati Bora</th>
<th>AMC Javelin</th>
<th>Merc 450SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dodge Challenger</td>
<td>Ford Pantera</td>
<td>Merc 450SE</td>
<td>Merc 450SL</td>
<td>Merc 280C</td>
<td>Valiant</td>
<td>Hornet Sport</td>
<td>Merc 280</td>
</tr>
<tr>
<td>Pontiac Firebird</td>
<td>Ferrari Dino</td>
<td>Mazda RX4</td>
<td>Mazda RX4</td>
<td>Hornet 4 Driv</td>
<td>Volvo 142E</td>
<td>Toyota Corolla</td>
<td>Datsun 710</td>
</tr>
<tr>
<td>Merc 230</td>
<td>Merc 240D</td>
<td>Porsche 914</td>
<td>Fiat X1-9</td>
<td>Honda Civic</td>
<td>Lotus Europa</td>
<td>Fiat 128</td>
<td>Toyota Corolla</td>
</tr>
<tr>
<td>Cadillac Flee</td>
<td>Lincoln Conti</td>
<td>Camaro Z28</td>
<td>Duster 360</td>
<td>Chrysler Imp</td>
<td>Maserati Bor</td>
<td>AMC Javelin</td>
<td>Merc 450SL</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>------------</td>
<td>------------</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>Dodge Challenger</td>
<td>Ford Pantera</td>
<td>Merc 450SE</td>
<td>Merc 450SL</td>
<td>Merc 280C</td>
<td>Valiant</td>
<td>Hornet Sport</td>
<td>Merc 280</td>
</tr>
<tr>
<td>Pontiac Firebird</td>
<td>Ferrari Dino</td>
<td>Mazda RX4</td>
<td>Mazda RX4</td>
<td>Hornet 4 Drift</td>
<td>Volvo 142E</td>
<td>Toyota Corolla</td>
<td>Datsun 710</td>
</tr>
<tr>
<td>Merc 230</td>
<td>Merc 240D</td>
<td>Porsche 914</td>
<td>Fiat X1-9</td>
<td>Honda Civic</td>
<td>Lotus Europa</td>
<td>Fiat 128</td>
<td>Toyota Corolla</td>
</tr>
</tbody>
</table>
If growth (slope) is important, plot it directly.
The most important measurement should exploit the highest ranked encoding possible.

- Position along a common scale
- Position on identical but nonaligned scales
- Length
- **Angle** or Slope
- Area
- Volume or Density or Color saturation/hue
Observation: Pie charts are **ALWAYS** a mistake.
Piecharts are the information visualization equivalent of a roofing hammer to the frontal lobe. They have no place in the world of grownups, and occupy the same semiotic space as short pants, a runny nose, and chocolate smeared on one’s face. They are as professional as a pair of assless chaps.

Piecharts are the information visualization equivalent of a roofing hammer to the frontal lobe. They have no place in the world of grownups, and occupy the same semiotic space as short pants, a runny nose, and chocolate smeared on one’s face. They are as professional as a pair of assless chaps.

Who do you think did a better job in tonight’s debate?

Among Republicans:
- Donald Trump 47%
- Hillary Clinton 53%

Among Democrats:
<table>
<thead>
<tr>
<th>Among Republicans</th>
<th>Among Democrats</th>
</tr>
</thead>
<tbody>
<tr>
<td>Donald Trump 1%</td>
<td>Hillary Clinton 99%</td>
</tr>
</tbody>
</table>
Tables are preferable to graphics for many small data sets. A table is nearly always better than a dumb pie chart; the only thing worse than a pie chart is several of them, for then the viewer is asked to compared quantities located in spatial disarray both within and between pies... Given their low data-density and failure to order numbers along a visual dimension, **pie charts should never be used.**

-Edward Tufte, The Visual Display of Quantitative Information
Tables are preferable to graphics for many small data sets. A table is nearly always better than a dumb pie chart; the only thing worse than a pie chart is several of them, for then the viewer is asked to compare quantities located in spatial disarray both within and between pies... Given their low data-density and failure to order numbers along a visual dimension, pie charts should never be used.

-Edward Tufte, The Visual Display of Quantitative Information
Who do you think did a better job in tonight’s debate?

<table>
<thead>
<tr>
<th></th>
<th>Clinton</th>
<th>Trump</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among Democrats</td>
<td>99%</td>
<td>1%</td>
</tr>
<tr>
<td>Among Republicans</td>
<td>53%</td>
<td>47%</td>
</tr>
</tbody>
</table>
All good pie charts are jokes...
The most important measurement should exploit the highest ranked encoding possible.

- Position along a common scale
- **Position on identical but nonaligned scales**
- Length
- Angle or Slope
- Area
- Volume or Density or Color saturation/hue
The most important measurement should exploit the highest ranked encoding possible.

- Position along a common scale
- Position on identical but nonaligned scales
- Length
- Angle or Slope
- Area
- Volume or Density or Color saturation/hue
Observation: Comparison is trivial on a common scale.
The chart shows a time series graph with two variables:

- **Value** on the y-axis.
- The x-axis represents time from 00:00 to 01:00.

Two lines are plotted:

- A red line labeled **cpu** exhibits a sharp peak at 00:30.
- A blue line labeled **latency** remains flat.

The graph is labeled **var** on the right side.
Today’s Learning Goals

• Appreciate the major elements of exploratory data analysis and why it is important to visualize data.

• Be conversant with data visualization best practices and understand how good visualizations optimize for the human visual system.

• Be able to generate informative graphical displays including scatterplots, histograms, bar graphs, boxplots, dendrograms and heatmaps and thereby gain exposure to the extensive graphical capabilities of R.

• Appreciate that you can build even more complex charts with ggplot and additional R packages such as rgl.
Different graphs for different purposes

**Exploratory graphs**: many images for a narrow audience (you!)
**Presentation graphs**: single image for a large audience
Roles of graphics in data analysis

- Graphs (& tables) are forms of communication:
  - What is the audience?
  - What is the message?

**Analysis graphs**: design to see patterns, trends, aid the process of data description, interpretation

**Presentation graphs**: design to attract attention, make a point, illustrate a conclusion

**Basic functions of data display**

Data Display → Analysis → Presentation

Primary Use
- Reconnaissance
- Exploration
- Diagnosis
- Model building
to Simulate
to Persuade
to Inform

Presentation Goal
- Exploratory (for you!)
- Info for others, publications & sharing etc.
Core R Graph Types
The R Painters Model

Side-Note: “Red and green should never be seen”
Core Graph Types

- Local options to change a specific plot
- Global options to affect all graphs
Common Options

- **Axis scales**
  - `xlim c(min,max)`
  - `ylim c(min,max)`

- **Axis labels**
  - `xlab(text)`
  - `ylab(text)`

- **Plot titles**
  - `main(text)`
  - `sub(text)`

- **Plot characters**
  - `pch(number)`
  - `cex(number)`

- Local options to change a specific plot
- Global options to affect all graphs
Plot Characters

cex sizes

Plot Characters

0.25  0.5  0.75  1   2   4   8

4    9    14   19   24

3    8    13   18   23

2    7    12   17   22

1    6    11   16   21

0    5    10   15   20
Plot Type Specific Options
Plot (scatterplots and line graphs)

- Input: Almost anything. 2 x Vectors
- Output: Nothing
- Options:
  - `type` `l=line`, `p=point`, `b=line+point`
  - `lwd` line width (thickness)
  - `lty` line type (1=solid, 2=dashed, 3=dotted etc.)

```
plot( c(1:10)^2, typ="b", lwd=4, lty=3 )
```
Barplot (bar graphs)

- **Input:** Vector (single) or Matrix (stack or group)
- **Output:** Bar centre positions
- **Options:**
  - `names.arg` Bar labels (if not from data)
  - `horiz=TRUE` Plot horizontally
  - `beside=TRUE` Plot multiple series as a group not stacked

```r
code: barplot(VADeaths, beside = TRUE)
```

![Bar Chart](image1)

![Stacked Bar Chart](image2)

![Grouped Bar Chart](image3)
Hist (histograms)

- **Input:** Vector
- **Output:** Summary of binned data
- **Options:**
  - `breaks` Number or limits of bins
  - `probability` Y axis is probability, not freq
  - `labels` Per bin text labels

```r
hist( c( rnorm(1000,0), rnorm(1000,4) ), breaks=20 )
```
Boxplot

- **Input:** Vector, List or formula (\texttt{data~factor})
- **Output:** Summary of the boxplot parameters
- **Options:**
  - \texttt{range} Sensitivity of whiskers
  - \texttt{varwidth} Width represents total observations
  - \texttt{horizontal} Plot horizontally

\begin{verbatim}
boxplot( cbind( rnorm(1000,0), rnorm(1000,4) ) )
\end{verbatim}
Controlling plot area options with `par`
Par

• The \texttt{par} function controls global parameters affecting all plots in the current plot area

• Changes affect all subsequent plots

• Many \texttt{par} options can also be passed to individual plots
Par examples

• Reading current value
  – `par()$cex`

• Setting a value
  – `par(cex=1.5) -> old.par`

• Restoring a value
  – `par(old.par)`
  – `dev.off()`
Par options

• Margins
  – \texttt{mai} (set margins in inches)
  – \texttt{mar} (set margins in number of lines)
  – \texttt{mex} (set lines per inch)
  – 4 element vector (bottom, left, top, right)

• Warning
  – Error in plot.new() : figure margins too large
Par options

• Fonts and labels
  – cex - global char expansion
    • cex.axis
    • cex.lab
    • cex.main
    • cex.sub
Par options

• Font style
  – $\text{font}(\text{font.axis, font.main, font.sub, font.lab})$
    • 1 = Plain text
    • 2 = Bold text
    • 3 = Italic text
    • 4 = Bold italic text
  – $\text{las}$ (label orientation)
    • 0 = Parallel to axis
    • 1 = Horizontal
    • 2 = Perpendicular
    • 3 = Vertical
Par options

- Multi-panel
  - `mfrow(rows, cols)`
  - Not supported by some packages
Exercise 1
Using Color
Specifying colors

- **Hexadecimal strings**
  - `#FF0000` (red)
  - `#0000FF` (blue)
  - `#CC00CC` (purple)

- **Controlled names**
  - "red" "green" etc.
  - `colors()`
Built in color schemes

• Functions to generate colors
• Pass in number of colors to make
• Functions:
  – rainbow
  – heat.colors
  – cm.colors
  – terrain.colors
  – topo.colors
Color Packages

• Color Brewer
  – Set of pre-defined, optimized palettes
  – library(RColorBrewer)
  – brewer.pal(no colours, palette)

• ColorRamps
  – Create smooth palettes for ramped color
  – Generates a function to make actual color vectors
  – colorRampPalette(c(“red”,“white”,“blue”))
  – colorRampPalette(c(“red”,“white”,“blue”))(5)
Color Packages

- **Colorspace**
  - `library(colorspace)`
  - `choose.palette()`
Applying Color to Plots

• Vector of colors passed to the `col` parameter

• Vector of factors used to divide the data
  – Colors taken from pallete
  – Can read or set using pallete function
    • `palette()`
    • `palette(brewer.pal(9,"Set1"))`
    • Ordered by levels of factor vector
Dynamic use of color

- Coloring by density
  - Pass data and palette to `densCols`
  - Vector of colors returned

- Coloring by value
  - Need function to map values to colors
map.colors <- function(value, range, palette) {
    proportion <- (value-range[1])/(range[2]-range[1])
    index <- round(((length(palette)-1)*proportion)+1)
    return(palette[index])
}
Exercise 2
Plot Overlays
Exercise 3
Points

- **Input:** 2 Vectors (x and y positions)
- **Options:**
  - `pch`
  - `cex`
• **Input:**
  – Lines 2 vectors (x and y)
  – Arrows 4 vectors (x0,x1,y0,y1)
  – Abline Intercept and slope (or correlation object)

• **Options:**
  – lwd
  – angle (arrows)
• Input:
  – 2 vectors (x and y) for bounding region
• Options:
  – `col`
• Input:
  – Text, \( x, y \)

• Options:
  – \texttt{adj}  \ (x and y offsets)
  – \texttt{pos} \ (auto offset 1=below, 2=left, 3=above, 4=right)
Legend

• Input:
  – Position (x,y or “topright”, “bottomleft” etc)
  – Text labels

• Options:
  – **fill** (colours for shaded boxes)
  – **xpd=NA** (draw outside plot area)
Exercise 3