ggplot2

displaying spatial & temporal data
Credit

• ‘White’ slides are taken directly from Dianne Cook’s “IMAGe STATMOS Course on Visualization of Climate Data”

• http://streaming.stat.iastate.edu/~dicook/NCAR/

• Creative Commons Attribution-Noncommercial 3.0 United States License
ggplot2
ggplot2

“ggplot2 is a plotting system for R, based on the grammar of graphics, which tries to take the good parts of base and lattice graphics and none of the bad parts. It takes care of many of the fiddly details that make plotting a hassle (like drawing legends) as well as providing a powerful model of graphics that makes it easy to produce complex multi-layered graphics.”

http://ggplot2.org/
ggplot2

• “ease of use” vs. “customization”
  • user’s time is more important than customization
  • grammar rules reduce amount of small decisions

• made for fast iterations
Grammar

Underlying ggplot2 is a formal structure for defining a data plot

Provides enormous flexibility in producing data plots, how different plots are related

Elegant nature of plots is due to defaults based on good cognitive principles.

Based initially on Wilkinson (2001)’s grammar of graphics - “gg” stands for grammar of graphics
• “Geoms, short for geometric objects, describe the type of plot you will produce”
• 37 documented geoms.

http://docs.ggplot2.org/0.9.3.1/
geom statistics

- statistical transformations
- most common: identity
- common geoms: bin, boxplot, qq, quantile, smooth

http://docs.ggplot2.org/0.9.3.1/
geom layer(s)

- Parts:
  - data and aesthetic mapping,
  - a statistical transformation (stat)
  - a geometric object (geom)
  - a position adjustment

- typically display other columns within the same data

- can display completely new data

http://vita.had.co.nz/papers/layered-grammar.pdf
layer examples
ggplot2 objects

- ggplot2 plots are fully defined R objects
- have a special print method
- objects may be altered many times before printing
‘qplot’ function

• qplot(
  vars, # ‘x’ and/or ‘y’. Depends on geom
data,
  [geom = “point”,
   [other options]]
) + [more layers]
• \( p \leftarrow \text{qplot(Sepal.Length, Sepal.Width, data = iris)} \)

\( p \)
> str(p)

List of 9

$ data 'data.frame':  150 obs. of  5 variables:
  ..$ Sepal.Length: num [1:150] 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
  ..$ Sepal.Width : num [1:150] 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
  ..$ Petal.Length: num [1:150] 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
  ..$ Petal.Width : num [1:150] 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
  ..$ Species : Factor w/ 3 levels "setosa","versicolor",...
layers :List of 1
  ..$ :Classes 'proto', 'environment' <environment: 0x7fa9d48e8b48>
scales :Reference class 'Scales' [package "ggplot2"] with 1 fields
  ..$ scales: list()
  ..and 21 methods, of which 9 are possibly relevant:
  ..  add, clone, find, get_scales, has_scale, initialize, input, n,
  ..  non_position_scales
mapping :List of 2
  ..$ x: symbol Sepal.Length
  ..$ y: symbol Sepal.Width
theme : list()
coordinates:List of 1
  ..$ limits:List of 2
  ..  ..$ x: NULL
  ..  ..$ y: NULL
  ..- attr(*, "class")= chr [1:2] "cartesian" "coord"
facet :List of 1
  ..$ shrink: logi TRUE
  ..- attr(*, "class")= chr [1:2] "null" "facet"
plot_env :<environment: R_GlobalEnv>
labels :List of 2
  ..$ x: chr "Sepal.Length"
  ..$ y: chr "Sepal.Width"
  ..- attr(*, "class")= chr [1:2] "gg" "ggplot
• > str(p)
   List of 9
   ...
   $ layers : List of 1
   ...$ : Classes 'proto', 'environment'
   <environment: 0x7fa9d48e8b48>
   ...

• > p$ layers
   [[1]]
   geom_point:
   stat_identity:
   position_identity: (width = NULL, height = NULL)
print plot:
base vs. ggplot2

- dt <- dataBeingUsed
- ## base
  # start graphics device
  plot(x, y, data=dt)
  points(x2, y2, data=dt)
  lines(x, y, data=dt)
  lines(x2, y2, data=dt)
  #... more plot procedures
  # stop graphics device

- ## ggplot2
  p <- qplot(x, y, data = dt, geom = c("point", "line"))
  p <- p +
      geom_point(aes(x=x2,y=y2)) +
      geom_line(aes(x=x2,y=y2))
  #... more plot procedures
  # start graphics device
  p
  # stop graphics device
ggplot2: spatial & temporal data
About the NASA data

Measurements recorded by NASA using remote sensing on temperature, pressure, ozone, clouds over Central America 1995-2000.

For ASA Data Expo 2006

http://stat-computing.org/dataexpo/2006/
## Checking the data

### Numerical summaries

```r
> summary(nasa)

<table>
<thead>
<tr>
<th></th>
<th>time</th>
<th>y</th>
<th>x</th>
<th>lat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>-21.20</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>18.75</td>
<td>6.75</td>
<td>6.75</td>
<td>-6.85</td>
</tr>
<tr>
<td>Median</td>
<td>36.50</td>
<td>12.50</td>
<td>12.50</td>
<td>7.50</td>
</tr>
<tr>
<td>Mean</td>
<td>36.50</td>
<td>12.50</td>
<td>12.50</td>
<td>7.50</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>54.25</td>
<td>18.25</td>
<td>18.25</td>
<td>21.85</td>
</tr>
<tr>
<td>Max.</td>
<td>72.00</td>
<td>24.00</td>
<td>24.00</td>
<td>36.20</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>long</th>
<th>date</th>
<th>cloudhigh</th>
<th>cloudlow</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>-113.8</td>
<td>1995-01-01: 576</td>
<td>0.0</td>
<td>0.50</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>-99.4</td>
<td>1995-02-01: 576</td>
<td>1.5</td>
<td>15.00</td>
</tr>
<tr>
<td>Median</td>
<td>-85.0</td>
<td>1995-03-01: 576</td>
<td>8.5</td>
<td>23.50</td>
</tr>
<tr>
<td>Mean</td>
<td>-85.0</td>
<td>1995-04-01: 576</td>
<td>12.0</td>
<td>26.24</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>-70.6</td>
<td>1995-05-01: 576</td>
<td>18.5</td>
<td>34.50</td>
</tr>
<tr>
<td>Max.</td>
<td>-56.2</td>
<td>1995-06-01: 576</td>
<td>62.5</td>
<td>84.50</td>
</tr>
<tr>
<td>(Other)</td>
<td></td>
<td>38016</td>
<td></td>
<td>NA's: 110.00</td>
</tr>
</tbody>
</table>
```

...
qplot(date, ozone, data = nasa, group = id, geom = "line", alpha = I(0.1))
Time trend

- Ozone is plotted against time (month and year) separately but overlapping for each spatial location.

- Seasonality is visible - but there is a double peak. We'd guess that this correspond to northern and southern latitude differences. How can we check?
qplot(date, ozone, data = nasa, geom = "line") + facet_grid(lat ~ long)
Space-time trend

- Ozone is plotted against time (month and year) separately for each spatial location.
- Seasonality is visible across the region, more so at the higher latitudes.

- This last plot took a minute to draw!
What is a map?

Set of **points** specifying latitude and longitude

**Polygon:** connect dots in correct order
What is a map?

Polygon: connect only the correct dots
What is a map?

Polygon: connect only the correct dots
states data

- `library(maps)`
- `states <- map_data("state")`
- `str(states)`

'data.frame': 15537 obs. of 6 variables:

$ long : num -87.5 -87.5 -87.5 -87.5 -87.6 ...
$ lat : num  30.4  30.4  30.4  30.3  30.3 ...
$ group : num  1 1 1 1 1 1 1 1 1 1 ...
$ order : int  1 2 3 4 5 6 7 8 9 10 ...
$ region : chr "alabama" "alabama" "alabama" "alabama" "alabama" ...
$ subregion: chr NA NA NA NA NA NA ...
> head(states)

<table>
<thead>
<tr>
<th>long</th>
<th>lat</th>
<th>group</th>
<th>order</th>
<th>region</th>
<th>subregion</th>
</tr>
</thead>
<tbody>
<tr>
<td>-87.46201</td>
<td>30.38968</td>
<td>1</td>
<td>1</td>
<td>alabama</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>-87.48493</td>
<td>30.37249</td>
<td>1</td>
<td>2</td>
<td>alabama</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>-87.52503</td>
<td>30.37249</td>
<td>1</td>
<td>3</td>
<td>alabama</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>-87.53076</td>
<td>30.33239</td>
<td>1</td>
<td>4</td>
<td>alabama</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>-87.57087</td>
<td>30.32665</td>
<td>1</td>
<td>5</td>
<td>alabama</td>
<td>&lt;NA&gt;</td>
</tr>
<tr>
<td>-87.58806</td>
<td>30.32665</td>
<td>1</td>
<td>6</td>
<td>alabama</td>
<td>&lt;NA&gt;</td>
</tr>
</tbody>
</table>
Maps in ggplot2

Geoms: polygon or path for filled polygons or outlines only

```r
> qplot(data = states, x = long, y=lat, 
      order = order, group = group, 
      geom = "path"/"polygon")
```
```r
qplot(long, lat, geom="point", data=states)

qplot(long, lat, geom="path", data=states, group=group, order = order)

qplot(long, lat, geom="polygon", data=states, group=group, order = order)
```
Making Maps look like Maps

```r
> theme(axis.title.x=element_blank(), # remove x and y label
axi.title.y=element_blank(),
axis.line=element_blank(), # no axis at the bottom and the left
axis.ticks=element_blank(), # don’t show ticks
axis.text.y = element_blank(), # no tick marks
axis.text.x = element_blank(),

panel.grid.minor=element_blank(), # don’t show any gridlines
panel.grid.major=element_blank(),
panel.background=element_blank(), # invisible background
panel.border = element_blank(), # no border around plot space
panel.margin = unit(0, "lines")
)
```
RGoogleMaps

Load a satellite image:
> library("ggmap")
> nasa_center <- c(lon=-85.0, lat=7.50)
> nasamap <- get_googlemap(center = nasa_center, zoom=4)
Map from URL: http://maps.googleapis.com/maps/api/staticmap?center=7.5,-85&zoom=4&size=%20640x640&maptype=terrain&sensor=false
Google Maps API Terms of Service: http://developers.google.com/maps/terms
> ggmap(nasamap)
Adding to satellite images

```r
> ggmap(nasamap) + geom_point(data=nasa,
  aes(x = long, y = lat))
```
Adding to satellite images

```r
> ggmap(nasamap) + geom_line(data=nasa.gly, aes(x=gx, y=gy, group = gid))
```
Questions?

Demo?
Links

• http://docs.ggplot2.org/

• http://streaming.stat.iastate.edu/~dicook/NCAR/

• http://vita.had.co.nz/papers/glyph-maps.pdf

• http://vita.had.co.nz/papers/layered-grammar.pdf

• Displaying time series, spatial and space-time data with R (not ggplot2) http://oscarperpinan.github.io/spacetime-vis/