# R graphics: overview, desiderata

Vince Carey
JSM 2002
stvjc@channing.harvard.edu

Overview

# R graphics: overview, desiderata

Vince Carey
JSM 2002
stvjc@channing.harvard.edu

Overview

- tentative basic definitions

# R graphics: overview, desiderata

Vince Carey
JSM 2002
stvjc@channing.harvard.edu

Overview

- tentative basic definitions
- principles of concrete stat viz

# R graphics: overview, desiderata

Vince Carey
JSM 2002
stvjc@channing.harvard.edu

Overview

- tentative basic definitions
- principles of concrete stat viz
- naive 3-tier architecture for R graphics

# R graphics: overview, desiderata

Vince Carey
JSM 2002
stvjc@channing.harvard.edu

Overview

- tentative basic definitions
- principles of concrete stat viz
- naive 3-tier architecture for R graphics
- some static examples

# R graphics: overview, desiderata

Vince Carey
JSM 2002
stvjc@channing.harvard.edu

Overview

- tentative basic definitions
- principles of concrete stat viz
- naive 3-tier architecture for R graphics
- some static examples
- generalized brushing via *tkrplot*

# R graphics: overview, desiderata

Vince Carey
JSM 2002
stvjc@channing.harvard.edu

Overview

- tentative basic definitions
- principles of concrete stat viz
- naive 3-tier architecture for R graphics
- some static examples
- generalized brushing via *tkrplot*
- positioning R for data mining

# Definition

**If**

Stat analysis = `Data` + $Story$

**then**

# Definition

**If**

Stat analysis = `Data` + $Story$

**then**

Stat graphics = `Data Graphics` + $Story\ Graphics$

# Definition

**If**

Stat analysis = `Data` + $Story$

**then**

Stat graphics = `Data Graphics` + $Story\ Graphics$

`Data Graphics` monographs by Tufte, Wilkinson, Chambers...

# Definition

**If**

Stat analysis = `Data` + *Story*

**then**

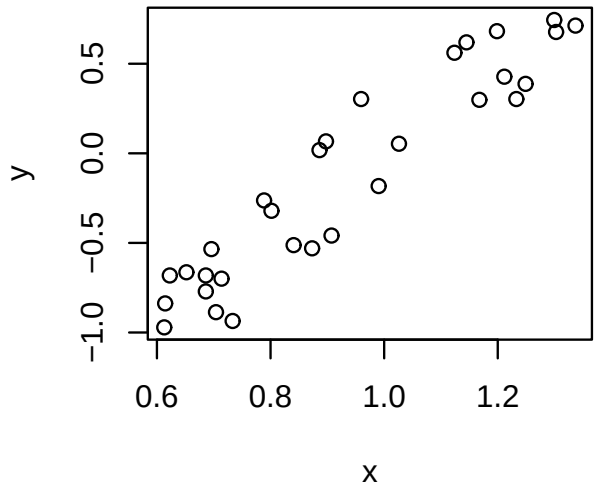
Stat graphics = `Data Graphics` + *Story Graphics*

`Data Graphics` monographs by Tufte, Wilkinson, Chambers...

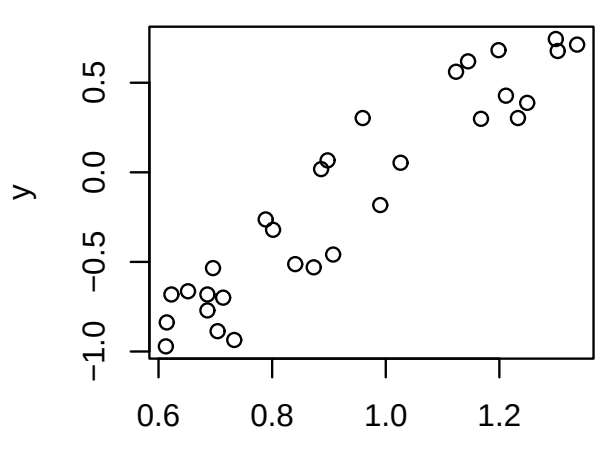
*Story Graphics* – ?

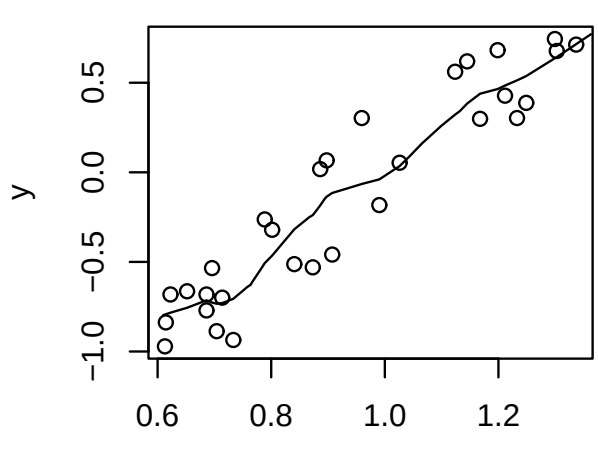
# a little data, a few stories

# a little data, a few stories

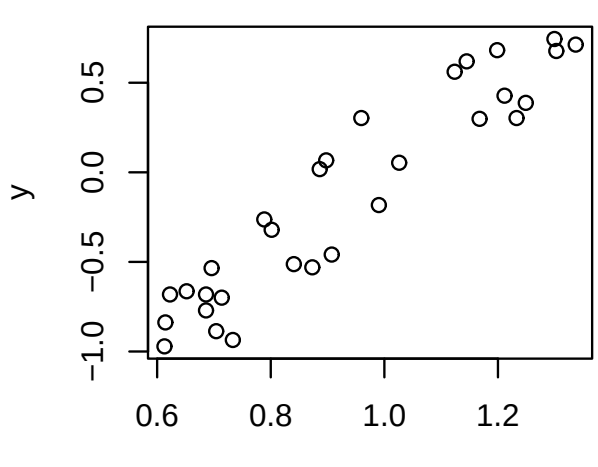


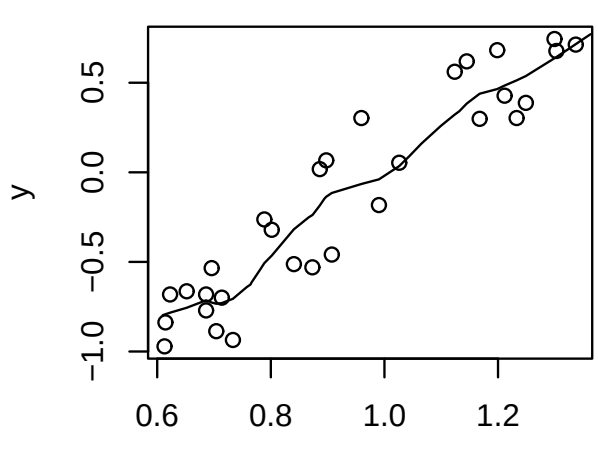
# a little data, a few stories

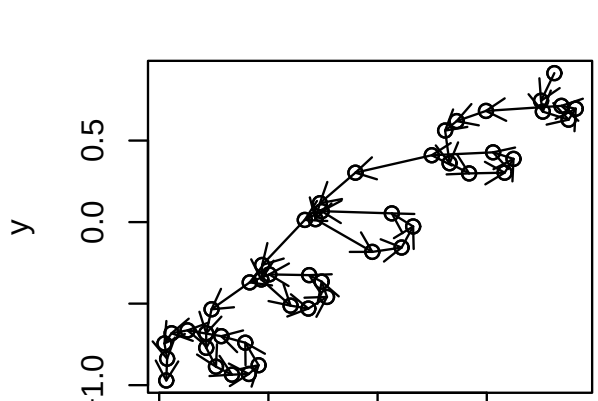


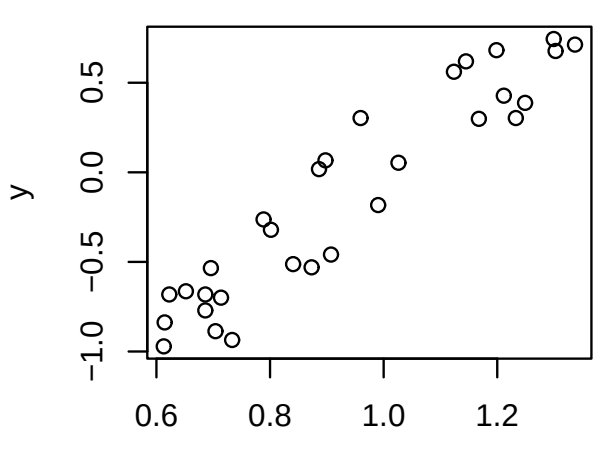
# a little data, a few stories

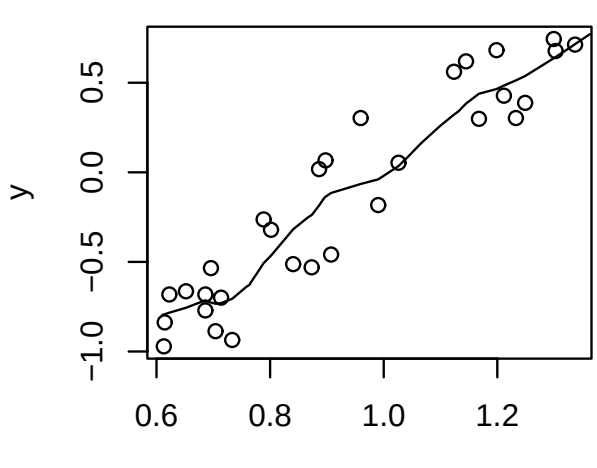


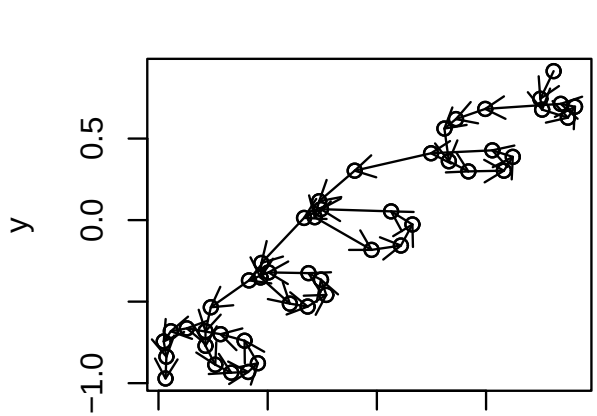



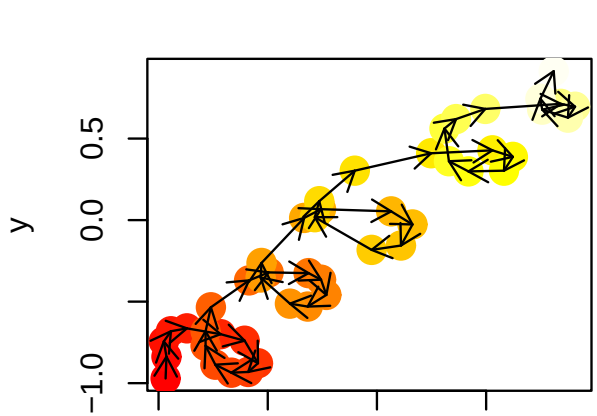



# a little data, a few stories

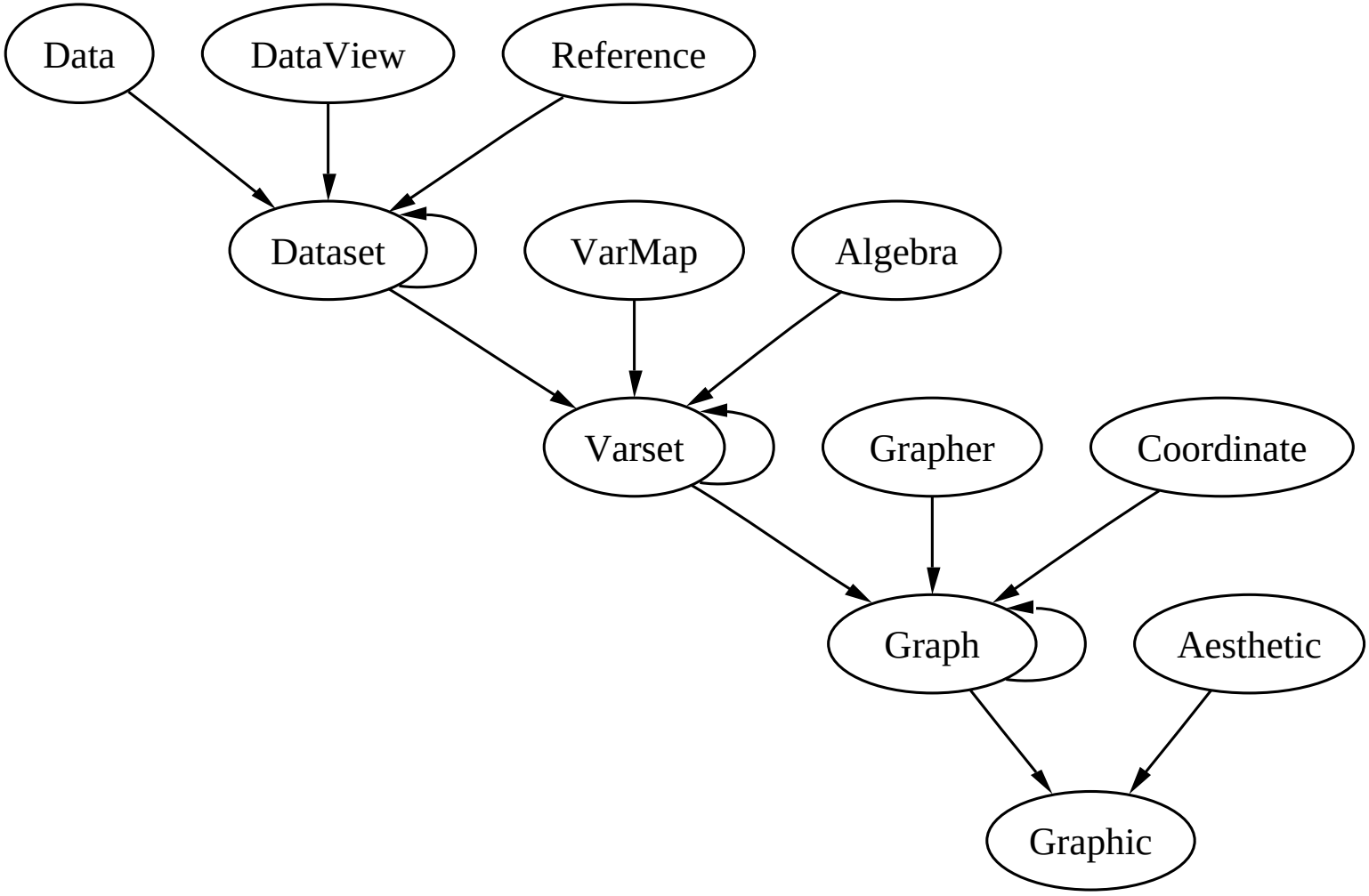


# Grammar sketch: Wilkinson The



# Statement of aims

# Statement of aims

- exploratory visualization: efficient navigation of story space for a set of data/hypotheses

# Statement of aims

- exploratory visualization: efficient navigation of story space for a set of data/hypotheses
- excellent illustration of stories once they are chosen

# Statement of aims

- exploratory visualization: efficient navigation of story space for a set of data/hypotheses
- excellent illustration of stories once they are chosen
- reckoning of the effects of choosing one story over another: can the effects of an early bandwidth setting on final inference be traced – or even just exhibited?

# Statement of aims

- exploratory visualization: efficient navigation of story space for a set of data/hypotheses
- excellent illustration of stories once they are chosen
- reckoning of the effects of choosing one story over another: can the effects of an early bandwidth setting on final inference be traced – or even just exhibited?

# Principles of concrete statViz

# Principles of concrete statViz

- Data are sacred

# Principles of concrete statViz

- Data are sacred (but jitterable?)

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz
- Use good statistical analysis practices

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz
- Use good statistical analysis practices
  - Story space restrictions are to be minimized (rich family of analysis tools)

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz
- Use good statistical analysis practices
  - Story space restrictions are to be minimized (rich family of analysis tools)
  - Transparency of data manipulations/selections (open auditing)

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz
- Use good statistical analysis practices
  - Story space restrictions are to be minimized (rich family of analysis tools)
  - Transparency of data manipulations/selections (open auditing)
- Use good computer science practices

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz
- Use good statistical analysis practices
  - Story space restrictions are to be minimized (rich family of analysis tools)
  - Transparency of data manipulations/selections (open auditing)
- Use good computer science practices
  - reuse: functions, stylesheets, packages

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz
- Use good statistical analysis practices
  - Story space restrictions are to be minimized (rich family of analysis tools)
  - Transparency of data manipulations/selections (open auditing)
- Use good computer science practices
  - reuse: functions, stylesheets, packages
  - environment-independence: interoperability, durability

# Principles of concrete statViz

- Data are sacred (but jitterable?)
  - info on missingness, detection limits, interventions should be tightly coupled to viz
- Use good statistical analysis practices
  - Story space restrictions are to be minimized (rich family of analysis tools)
  - Transparency of data manipulations/selections (open auditing)
- Use good computer science practices
  - reuse: functions, stylesheets, packages
  - environment-independence: interoperability, durability
  - self-documenting objects preferred

# Naive 3-tier architecture

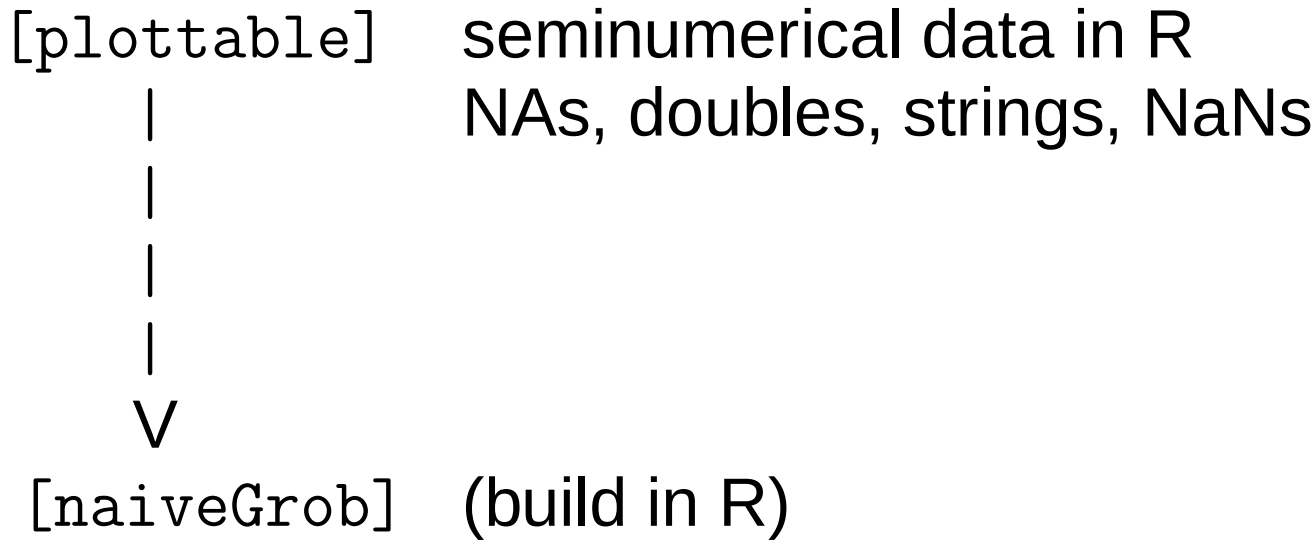
# Naive 3-tier architecture

```
[plottable]
    |
    |
```

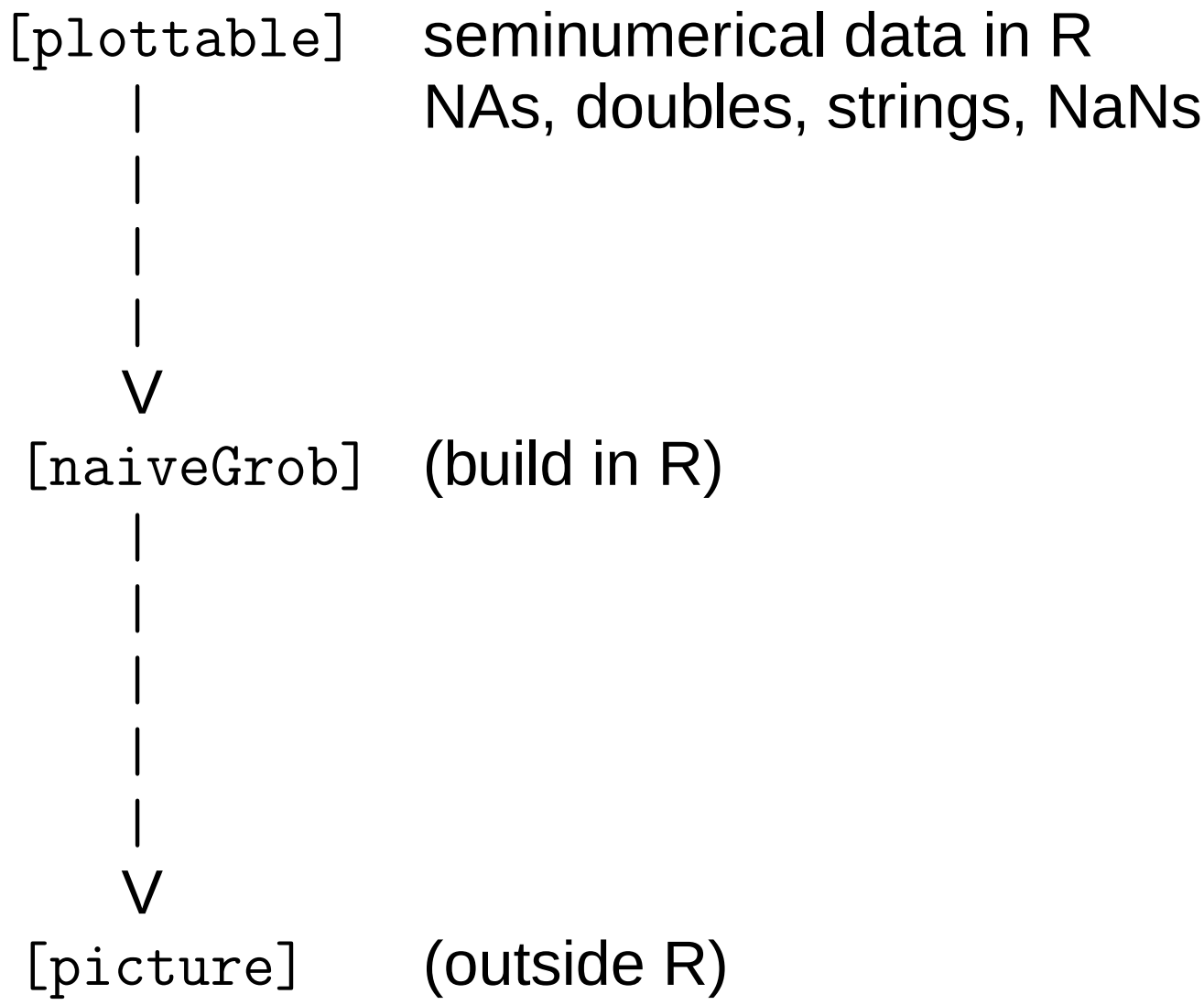
seminumerical data in R

NAs, doubles, strings, NaNs

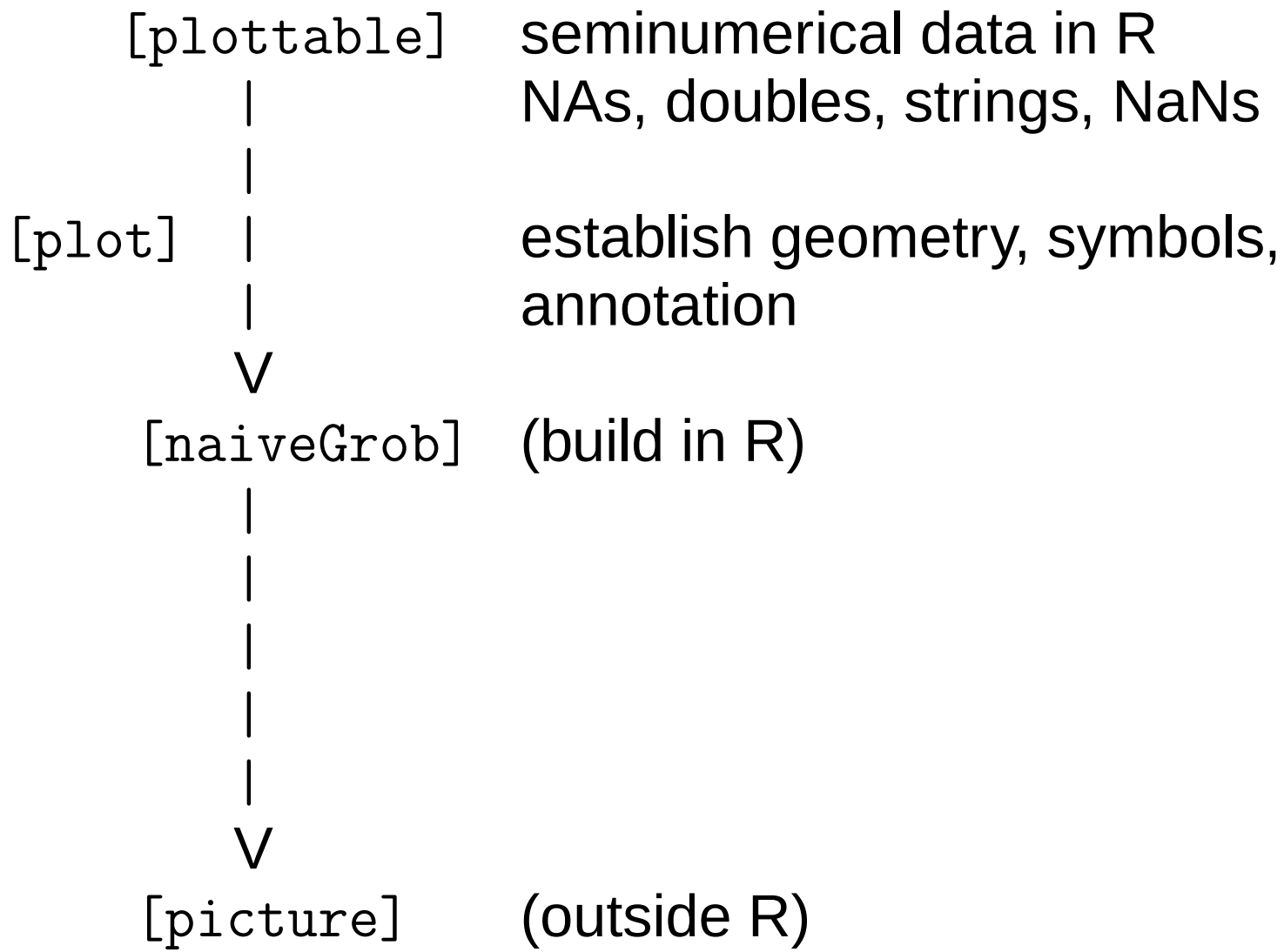
# Naive 3-tier architecture



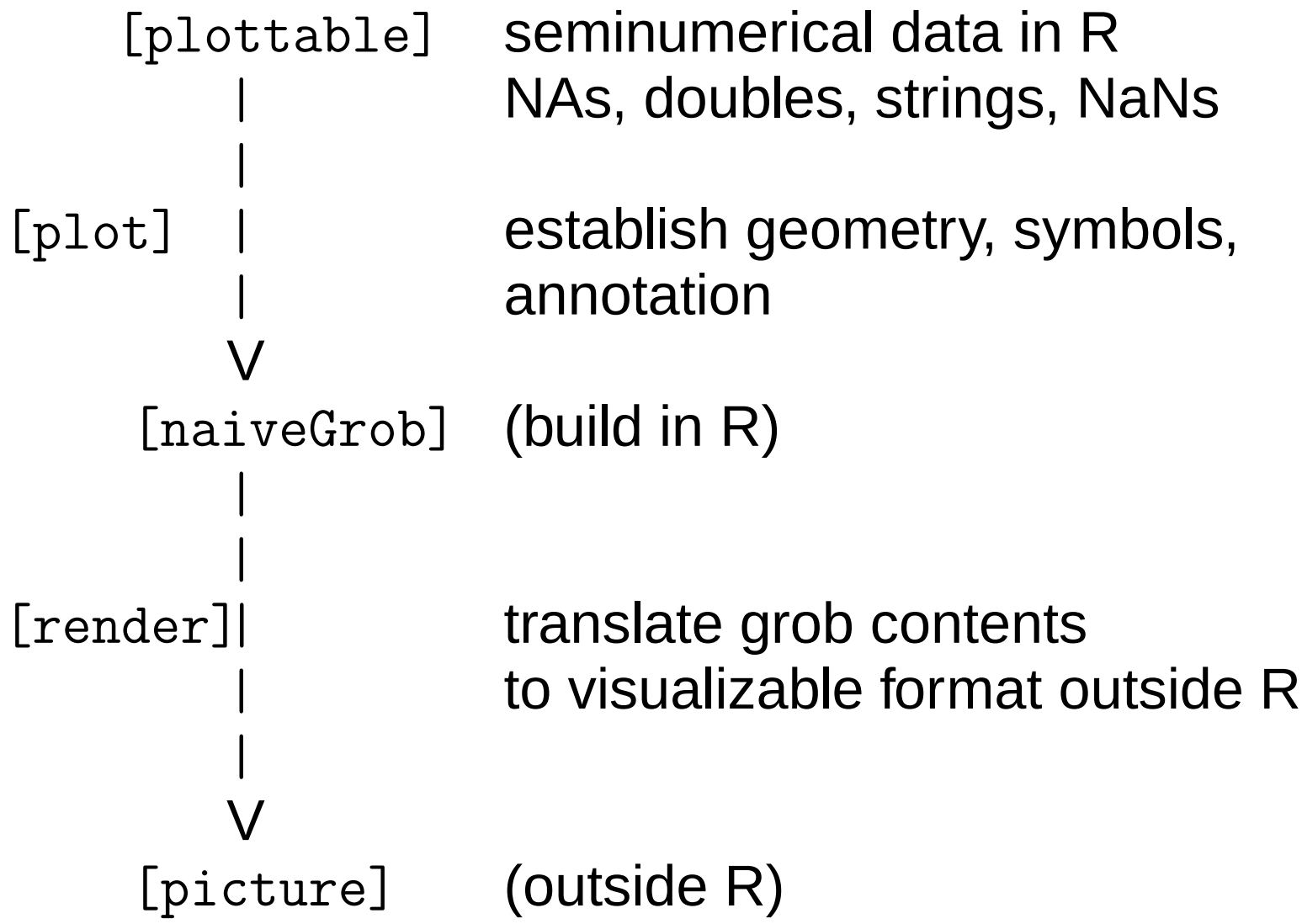
# Naive 3-tier architecture



# Naive 3-tier architecture



# Naive 3-tier architecture



# render targets: species of picture

# render targets: species of picture

| driver | output | character |
|---|---|---|
| x11 | bitmap | (static/interrogate/ augment) |

# render targets: species of picture

| driver | output | character |
|---|---|---|
| x11 | bitmap | (static/interrogate/ augment) |
| postscript, pdf | program | (static+) |

# render targets: species of picture

| driver | output | character |
|---|---|---|
| x11 | bitmap | (static/interrogate/ augment) |
| postscript, pdf | program | (static+) |
| jpeg,png | | (static) |

# render targets: species of picture

| driver | output | character |
|---|---|---|
| x11 | bitmap | (static/interrogate/ augment) |
| postscript, pdf | program | (static+) |
| jpeg,png | | (static) |
| xfig | program | (static+) |

# render targets: species of picture

| driver | output | character |
|---|---|---|
| x11 | bitmap | (static/interrogate/ augment) |
| postscript, pdf | program | (static+) |
| jpeg,png | | (static) |
| xfig | program | (static+) |
| java | object | (mutable) |

# render targets: species of picture

| driver | output | character |
|---|---|---|
| x11 | bitmap | (static/interrogate/ augment) |
| postscript, pdf | program | (static+) |
| jpeg,png | | (static) |
| xfig | program | (static+) |
| java | object | (mutable) |
| tkrplot, gtk | widget | (event-looped) |

# render targets: species of picture

| driver | output | character |
|---|---|---|
| x11 | bitmap | (static/interrogate/ augment) |
| postscript, pdf | program | (static+) |
| jpeg,png | | (static) |
| xfig | program | (static+) |
| java | object | (mutable) |
| tkrplot, gtk | widget | (event-looped) |
| SVG | XML document | (scriptable, XLST, ...) |

# Actual architecture

# Actual architecture

- grob development somewhat diffuse

# Actual architecture

- grob development somewhat diffuse
- burdens shared by function args, global options settings, themes

# Actual architecture

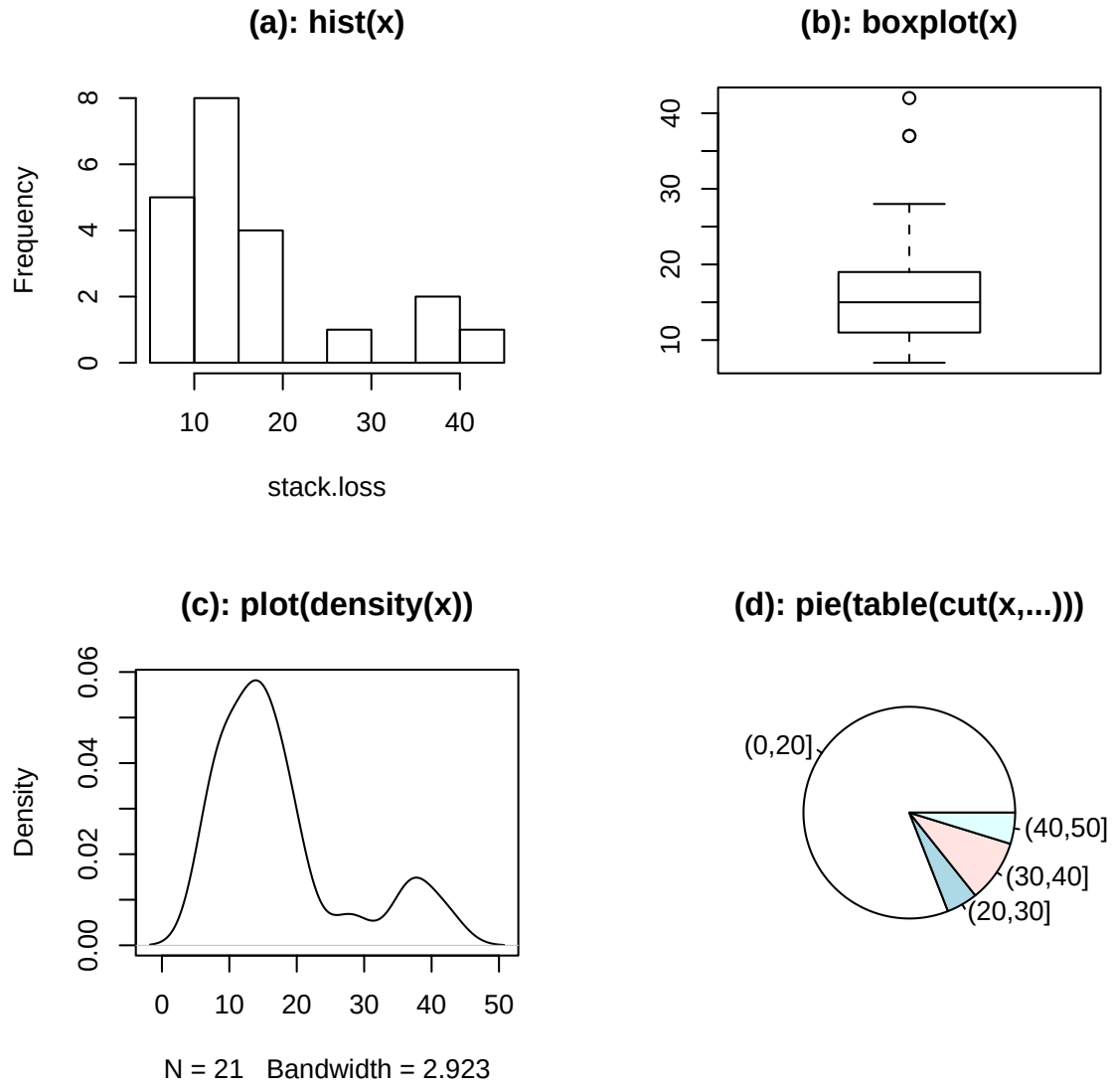
- grob development somewhat diffuse
- burdens shared by function args, global options settings, themes
- much is ephemeral: R program or .Rhistory can be the grob

# Actual architecture

- grob development somewhat diffuse
- burdens shared by function args, global options settings, themes
- much is ephemeral: R program or .Rhistory can be the grob
- outcome of rendering not always predictable, not uniform among targets

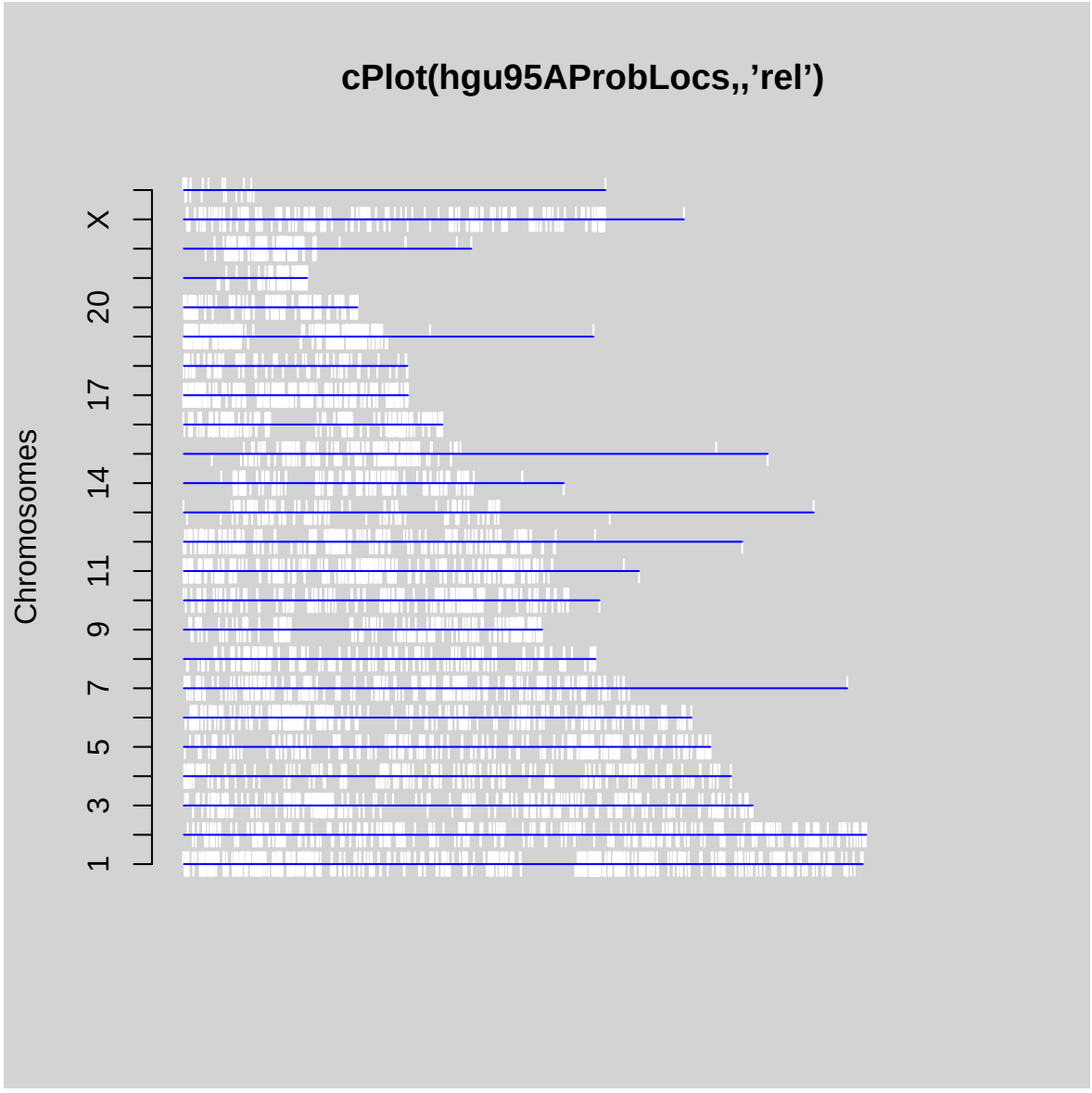
```
Loading required package: methods
Creating a new generic function for "plot" in pa
gelTools

Attaching package ‘gelTools’:


        The following object(s) are masked _by_ 

         plot


        The following object(s) are masked from 

         plot
```

# Univariate: `x <- stack.loss`

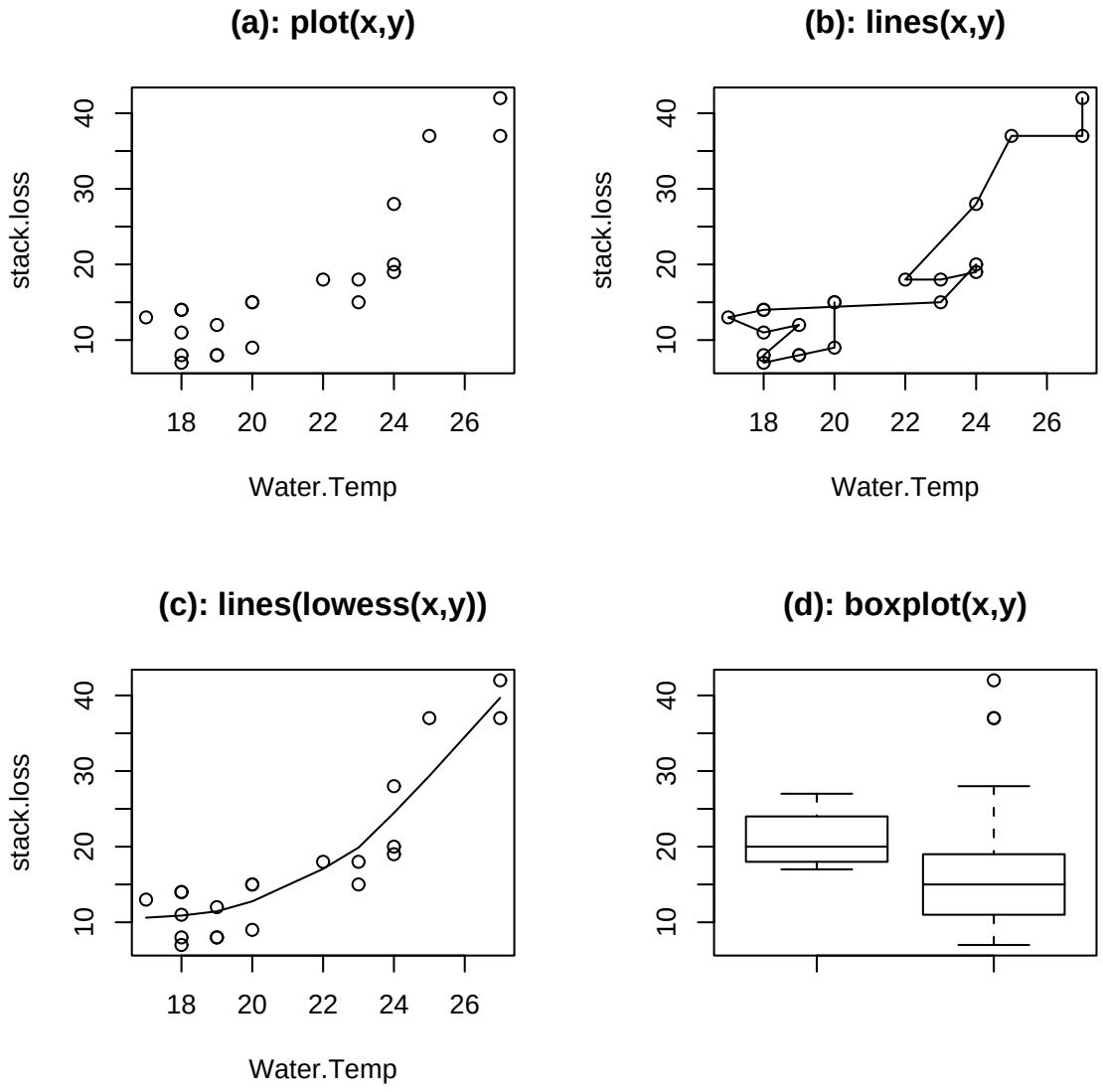


```
Loading required package: annotate
Loading required package: Biobase

Attaching package 'Biobase':


        The following object(s) are masked from p

         split

This is mgcv 0.8.1

Attaching package 'mgcv':


        The following object(s) are masked _by_

         s
```

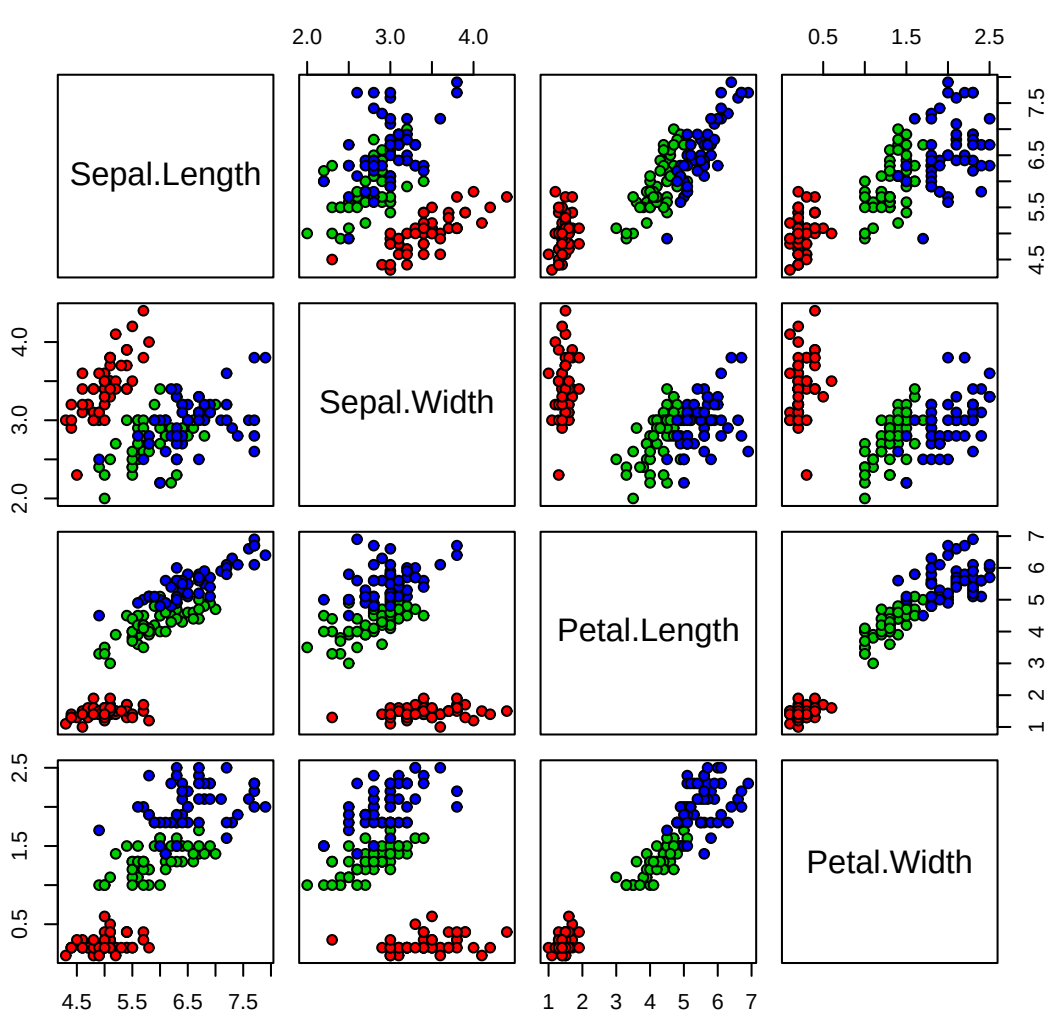
# Univ: U95Av2 probe locations



# Bivariate: y <- Water.temp

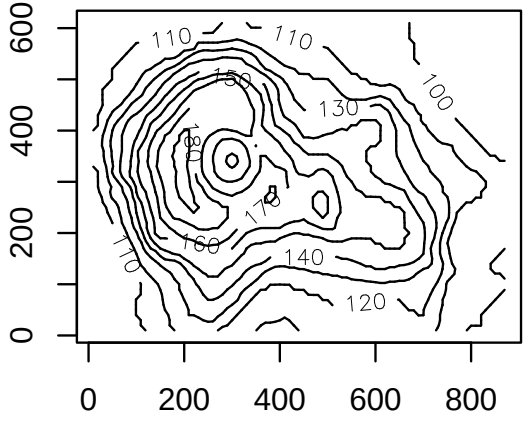


# Multi-bivariate

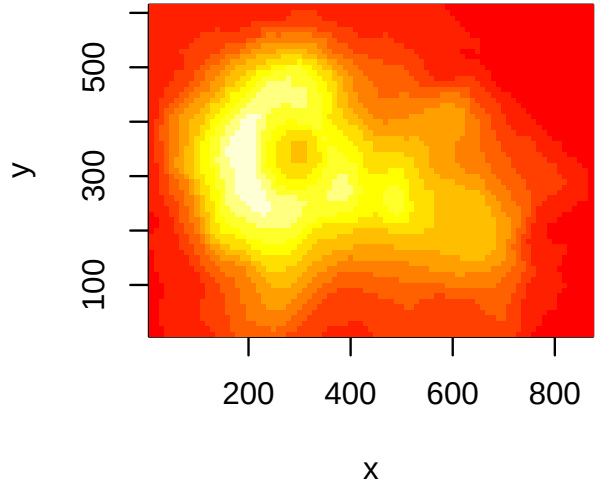


# Trivariate: volcano

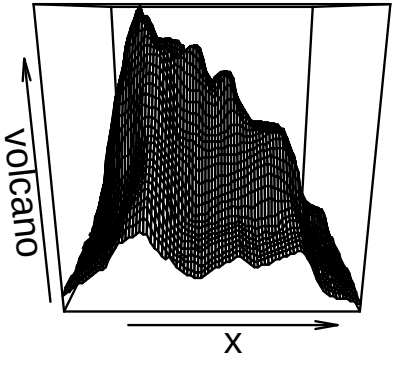
**(a): contour(x,y,z)**



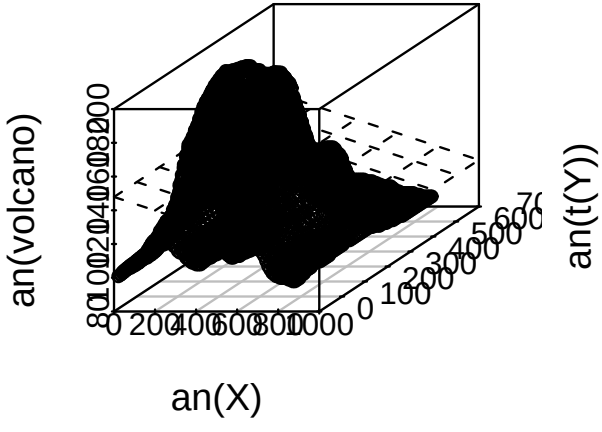

**(b): image(x,y,z)**




**(c): persp(x,y,z)**

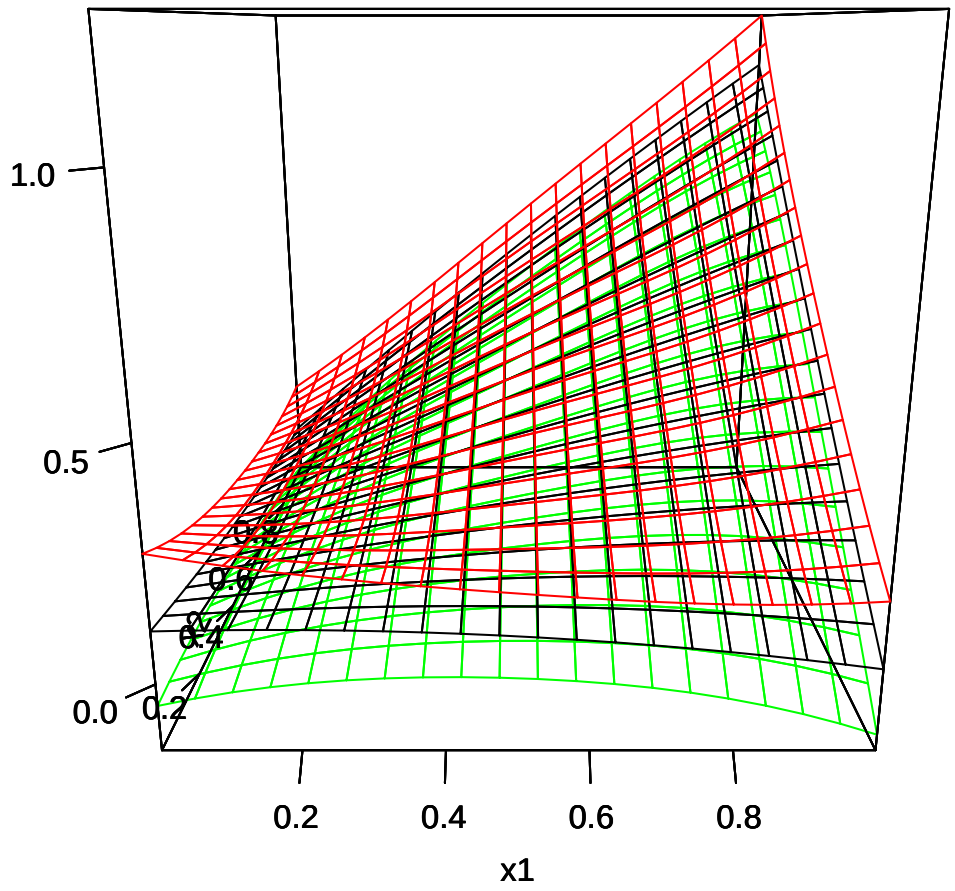



**(d): sctrplt3d(X,Y,z)+**



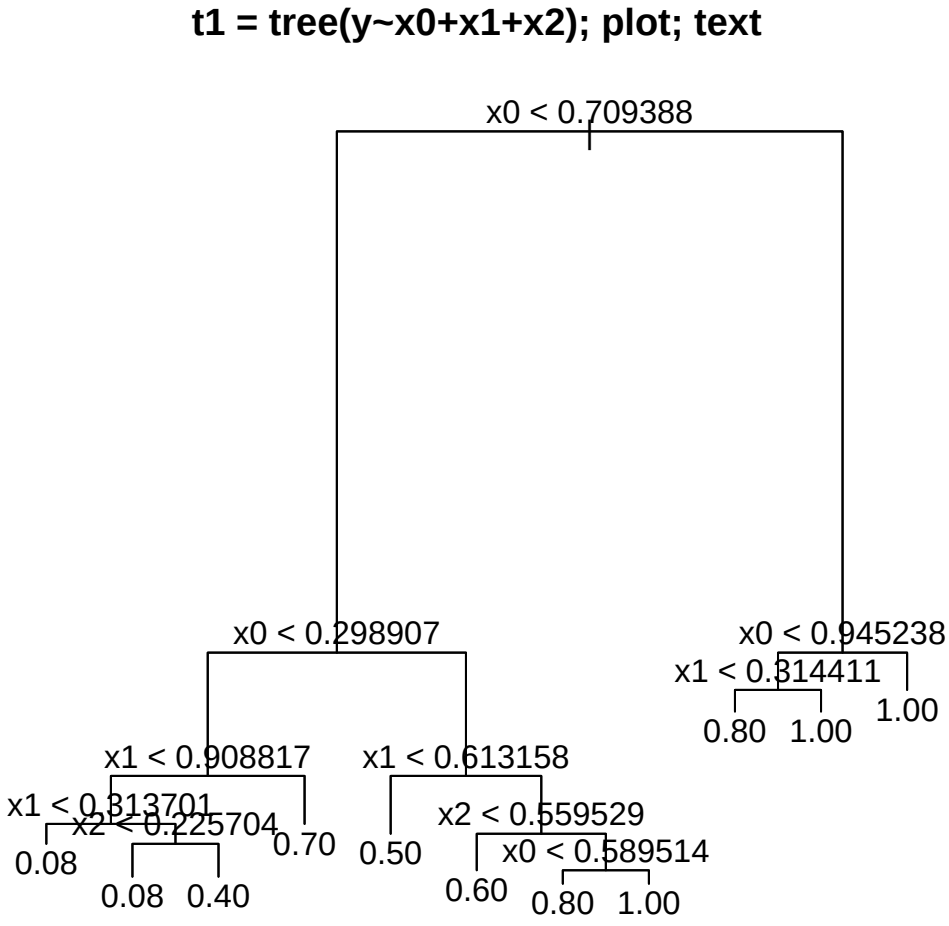
# Conditioning: gam via *mgcv*

$$y \approx x_0^2 + x_1 x_2 + U(-0.3, 0.3)$$

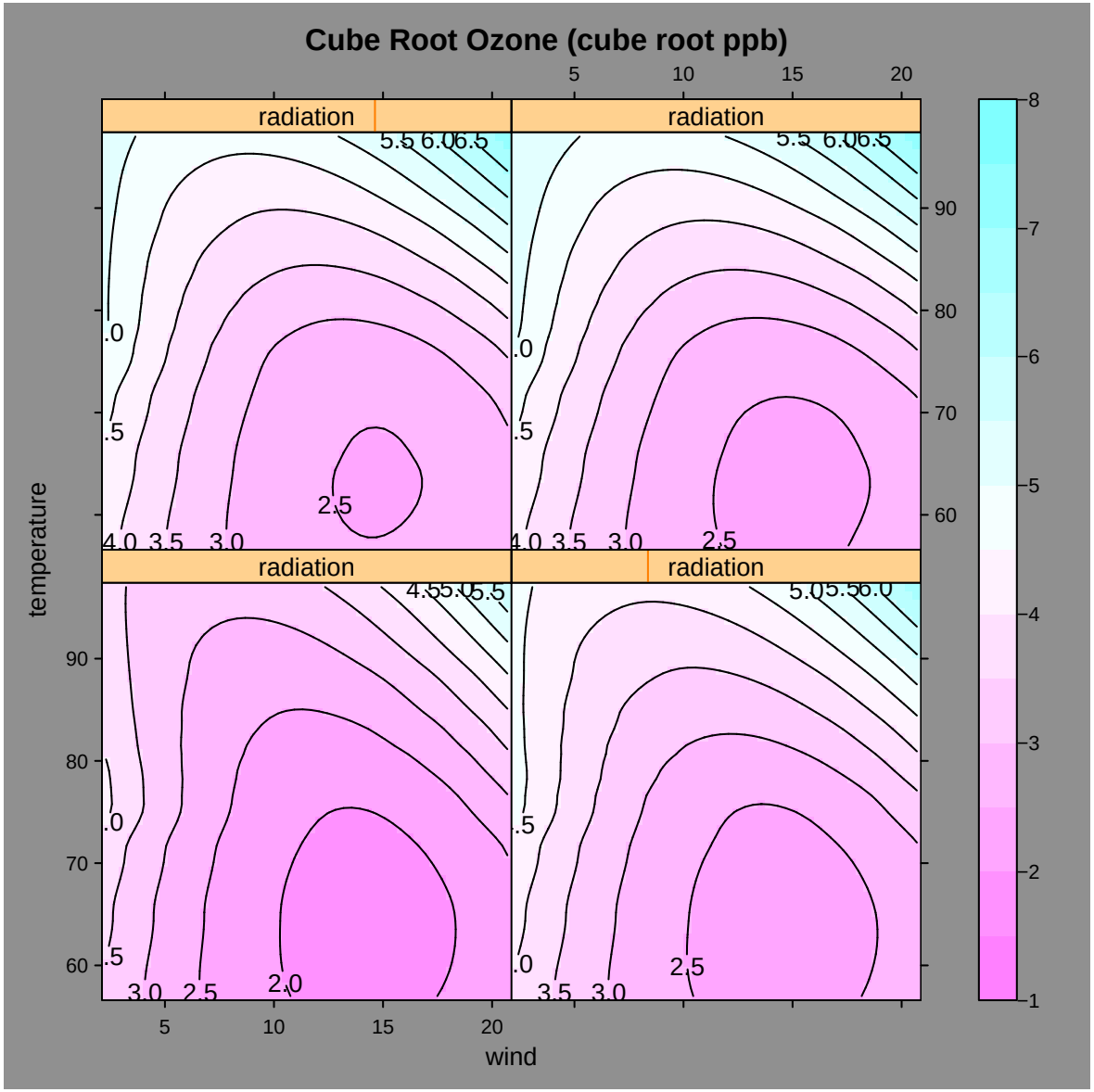



x0=0.25, red/green are +/−2 se

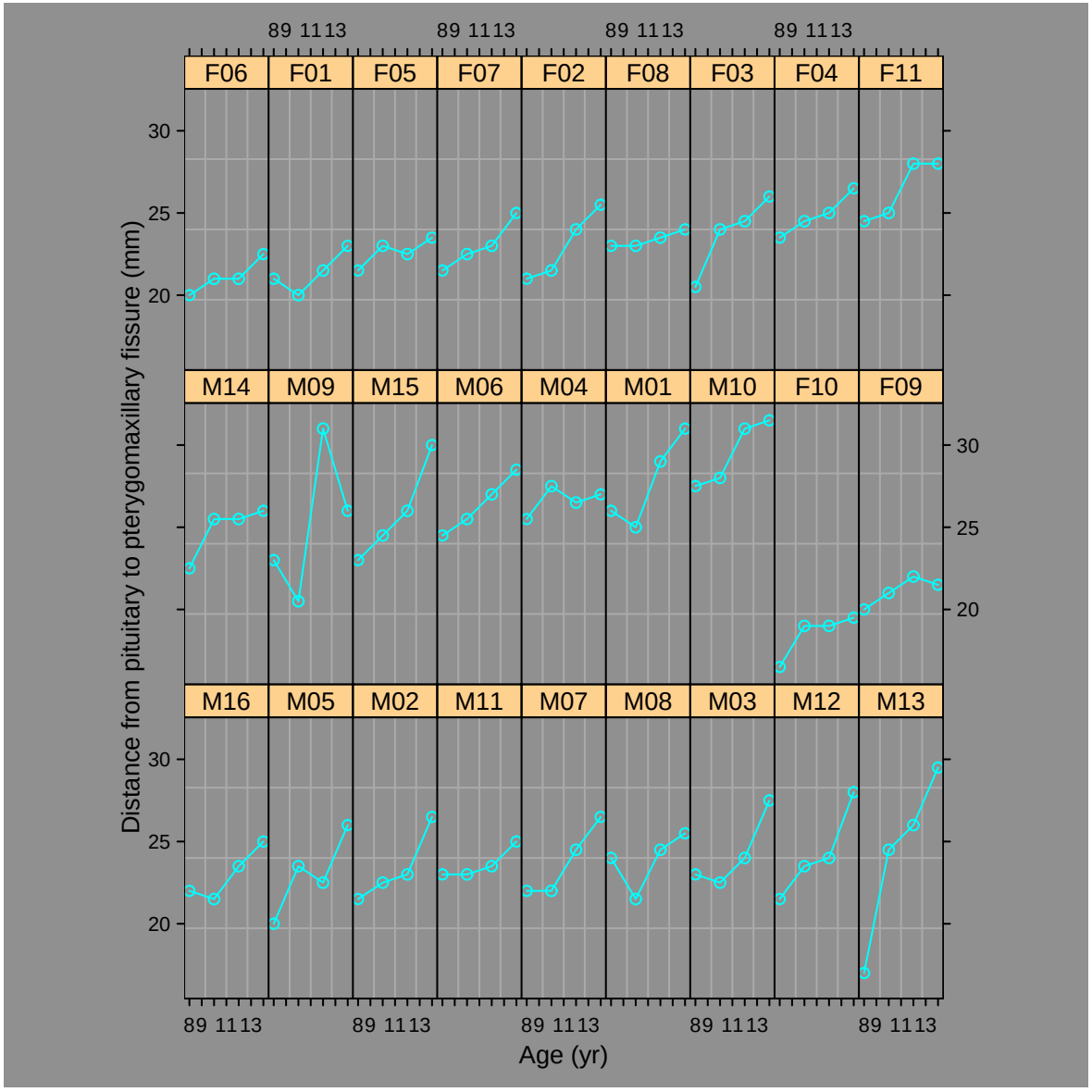
# Tree structured model



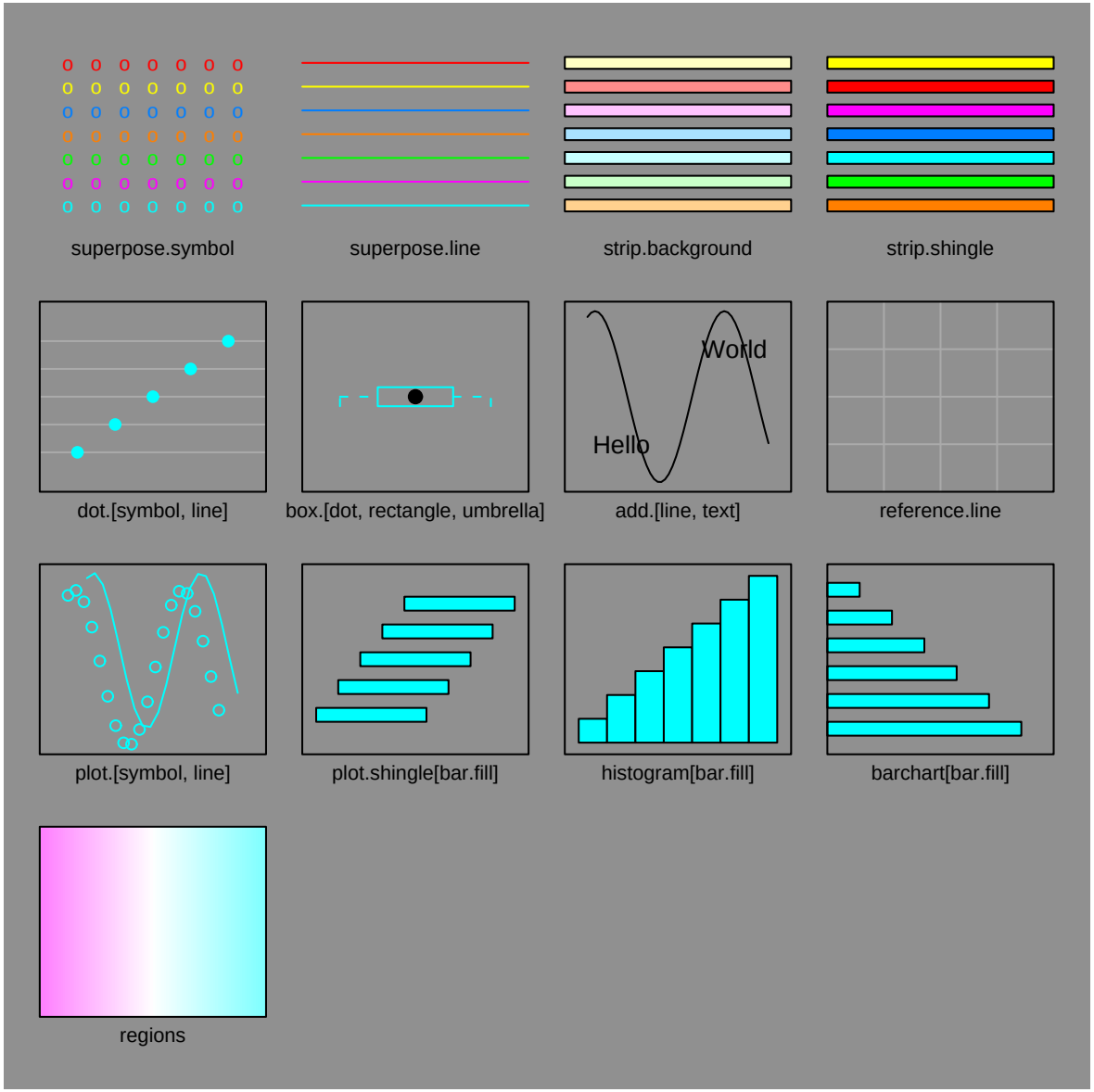
# lattice for 4-d data



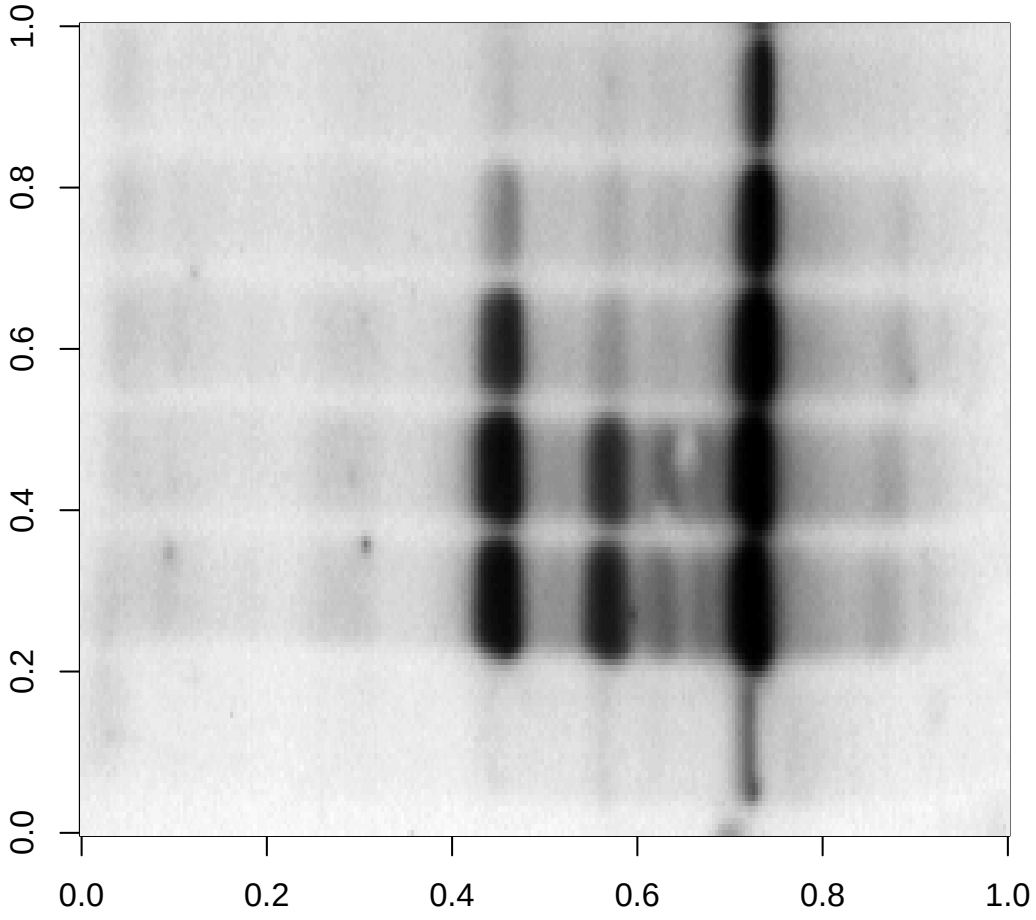
# lattice for longitudinal data



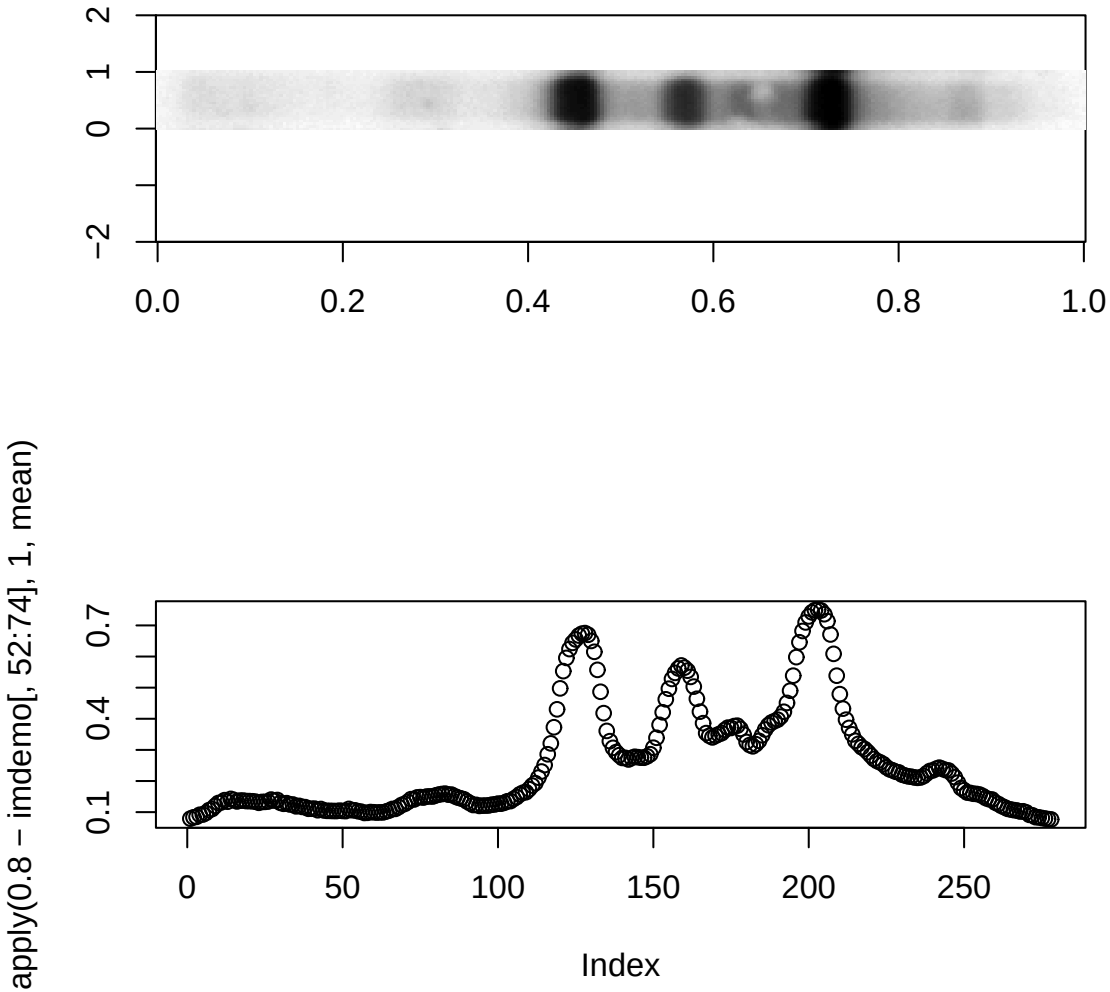
# lattice show.settings()



# handling images – pixmap



# profiling via pixel-level access



# Detour

to generalized brushing via $tkrplot$

# Not covered

- Superb annotation (`example(plotmath)`)
- document-graphics integration: $tools::Sweave$ generates this talk
- graphical objects via $grid$, the backbone of $lattice$
- new device drivers for Java and Gtk
- SVG

# Conclusions

# Conclusions

- $R$ $graphics$ embraces a number of scientific computing processes not ordinarily associated with data visualization:
  - close coupling of general modeling facilities to graph production facilities,
  - text-image integration,
  - browser-like interfaces to metadata...

# Conclusions

- $R$ $graphics$ embraces a number of scientific computing processes not ordinarily associated with data visualization:
  - close coupling of general modeling facilities to graph production facilities,
  - text-image integration,
  - browser-like interfaces to metadata...
- $R$ $graphics$ includes support of language and data structure facilities that permit thorough exploration of the hypothesis:data:inference complex.

# End

- As language and data structure facilities mature, the use of R for

# End

- As language and data structure facilities mature, the use of R for
  - *concept graphics*,

# End

- As language and data structure facilities mature, the use of R for
  - *concept graphics*,
  - *network diagrams and models*,

# End

- As language and data structure facilities mature, the use of R for
  - *concept graphics*,
  - *network diagrams and models*,
  - *metadata navigation* and

# End

- As language and data structure facilities mature, the use of R for
  - *concept graphics*,
  - *network diagrams and models*,
  - *metadata navigation* and
  - *data-user-interface design and deployment*

# End

- As language and data structure facilities mature, the use of R for
  - *concept graphics*,
  - *network diagrams and models*,
  - *metadata navigation* and
  - *data-user-interface design and deployment*

  will position R as a key tool in data mining platform construction and use.

# End

- As language and data structure facilities mature, the use of R for
  - *concept graphics*,
  - *network diagrams and models*,
  - *metadata navigation* and
  - *data-user-interface design and deployment*

  will position R as a key tool in data mining platform construction and use.

In the face of this potential, many practical problems (graph configurability, unpredictability of rendering, missing components [e.g., key placement, better tree rendering]) seem uncompelling.

Keep filing the bug reports ...