

BioTech 76

Creating Plots, Graph and Maps using R

Xin Tian, Ph.D., Mathematical Statistician

Office of Biostatistics Research
National Heart, lung and Blood Institute, NIH

Email: tianx@nhlbi.nih.gov

GitHub: <https://github.com/npmlbook/rplots>

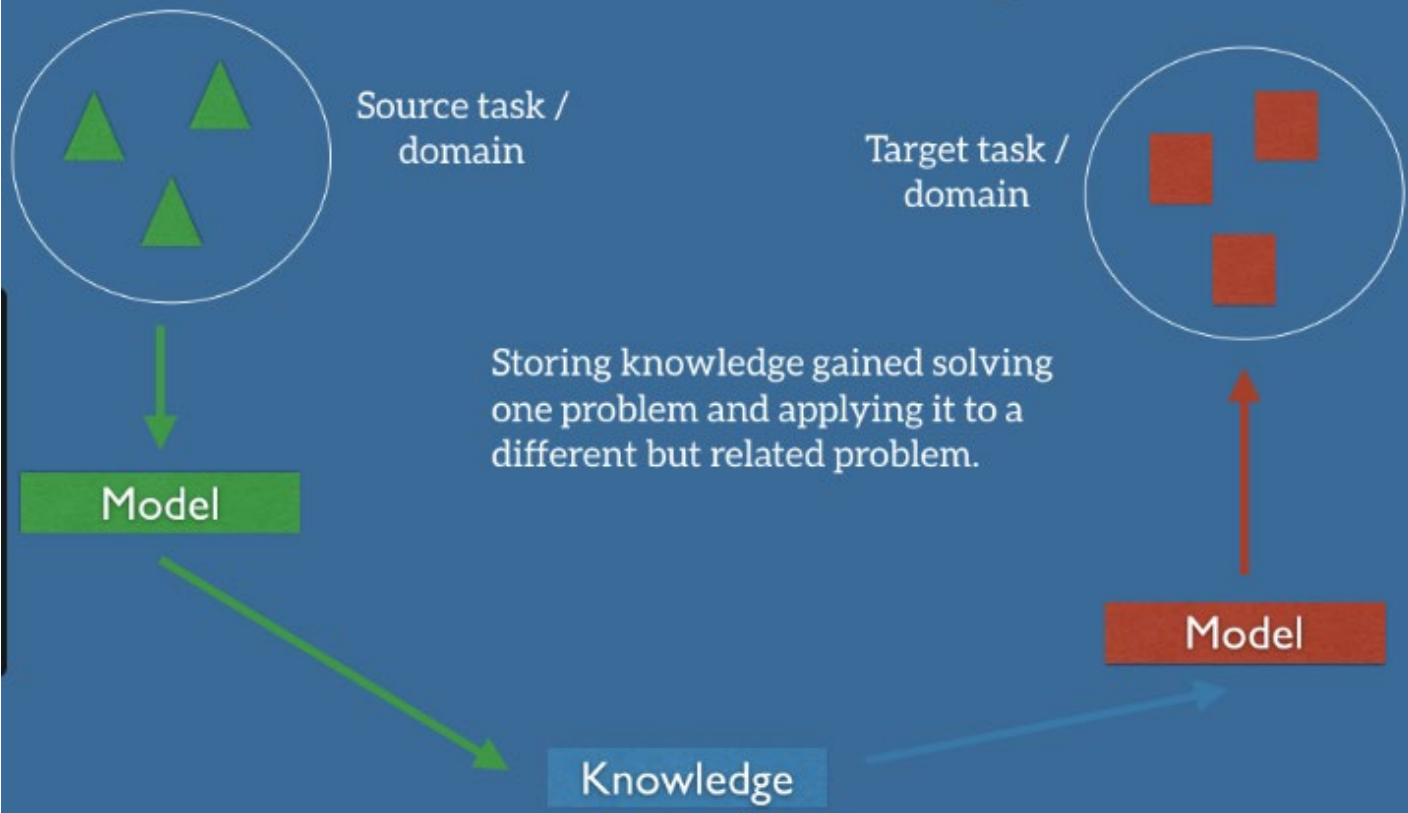
Agenda (breaks: 2*15min @10a, 3p & lunch)

- A : R basics
- B : Graphics systems in R
- **C : Using color in R**
- **D : R Graphics Devices – static plots**
- **E: R Graphics Devices – interactive plots**
- Q & A, hands-on and discuss your own research data
- **F: Working with Maps**
- **G: Network plots**
- **H: Specialty Figures**
- Q & A, hands on your own research data

Why we need plots/figures?

- *“A picture is worth a thousand words.”*
- *“Every picture tells a story.”*
- *“The graphical display of data is among the most powerful tools available for communicating medical research findings, given the increasing complexity of study designs and the mind’s preference for information conveyed in pictorial format.”* – [Stuart Pocock in ‘How to interpret figures in reports of clinical trials’, BMJ, 2008](#)

Transfer learning

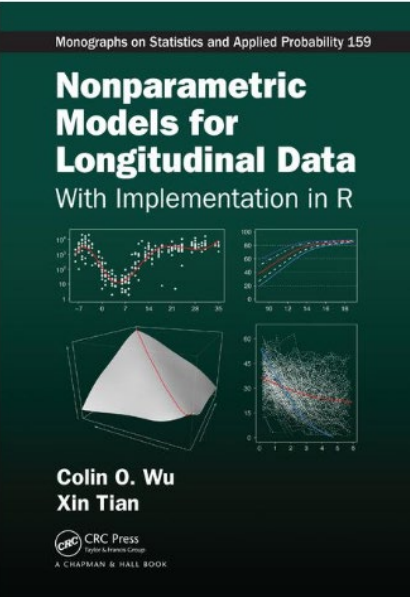


TRANSFER OF LEARNING

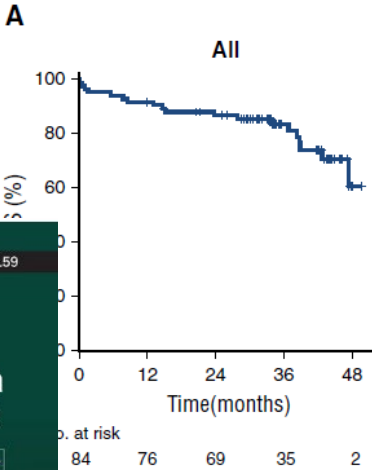


The application of skills, knowledge, and/or attitudes that were learned in one situation to another **learning** situation (Perkins, 1992)

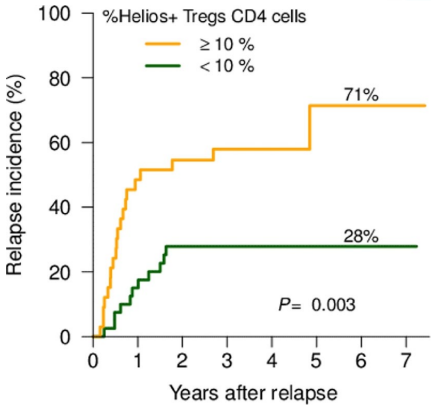
R Figures for My research



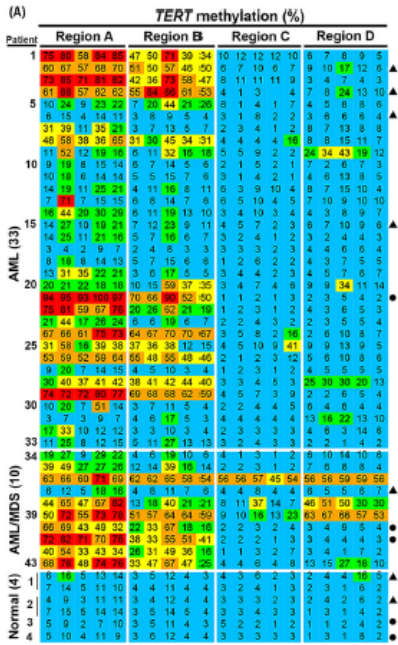
CRC press, 2018



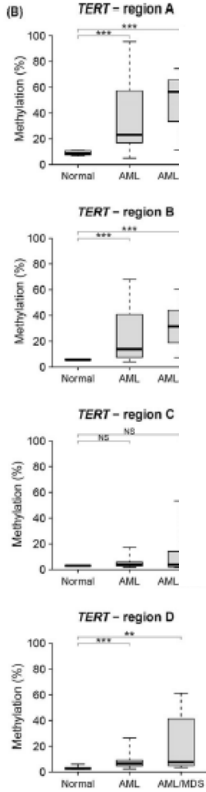
Blood, 2017



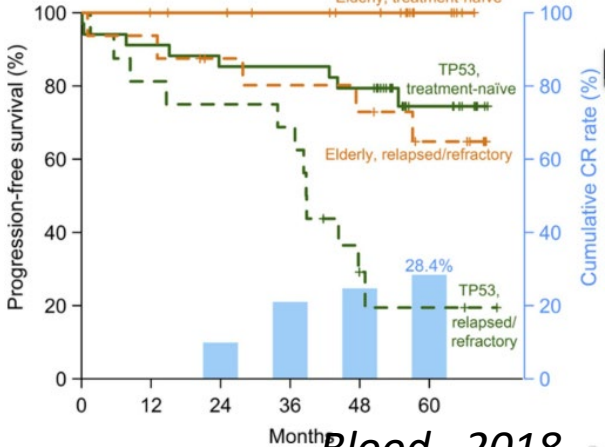
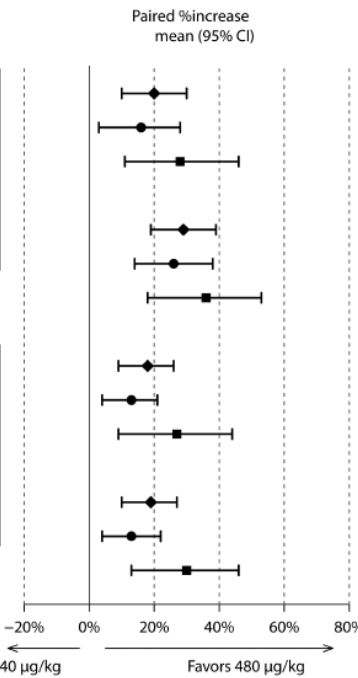
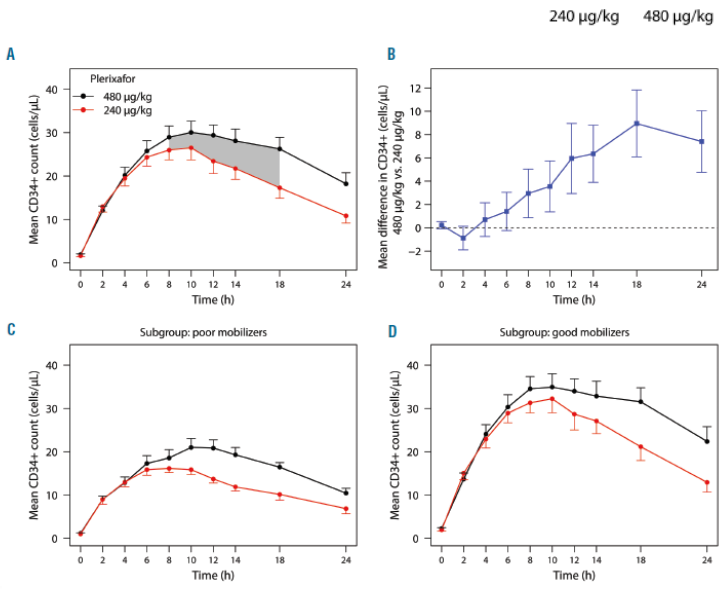
Biology of Blood & Marrow Transplant 2019



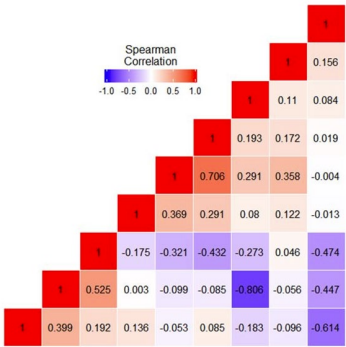
British J. Haematology 2016



Haematologica, 2017



Blood, 2018



BMT 2018

Grammar of R graphics = layers of elements/building blocks

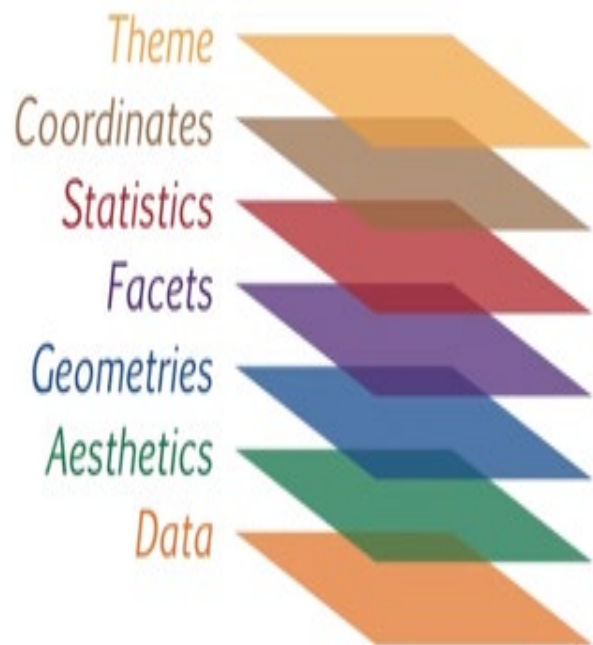
➤ Traditional R plots

- `plot()`; `hist()`; `boxplot()` ; `barplot()`
- `lines()`; `abline()`; `points()`
- `arrows()`; `segments()`
- `axis()`, `title()`, `legend()`
- `text()`, `mtext()`
- `grid()`, `box()`
- `par()`, `layout()`

➤ ggplot2 package



Diagram



| <i>Data</i> | <i>{variables of interest}</i> | | | | |
|--------------------|--------------------------------|------------------------------|------------------------------|------------------------------|---------------------------------------|
| <i>Aesthetics</i> | <i>x-axis</i> <i>y-axis</i> | <i>colour</i> <i>fill</i> | <i>size</i> <i>labels</i> | <i>alpha</i> <i>shape</i> | <i>line width</i> <i>line type</i> |
| <i>Geometries</i> | <i>point</i> | <i>line</i> | <i>histogram</i> | <i>bar</i> | <i>boxplot</i> |
| <i>Facets</i> | <i>columns</i> | <i>rows</i> | | | |
| <i>Statistics</i> | <i>binning</i> | <i>smoothing</i> | <i>descriptive</i> | <i>inferential</i> | |
| <i>Coordinates</i> | <i>cartesian</i> | <i>fixed</i> | <i>polar</i> | <i>limits</i> | |
| <i>Themes</i> | <i>non-data ink</i> | | | | |

Sec C 1-3: Using color in R- R can be colorful



Day2a_colo-r.R

- **Colors in R** : <http://research.stowers-institute.org/efg/R/Color/Chart>
 - > par(), arguments: col = "name" or number , col.axis=, col.lab=, col.main=, col.sub = & color/fill in ggplot2
 - > colors() : 657 build in colors
 - Use simple **numbers**, **co1=1 ~ 8** : default **palette()**
 - **use color names** (check the color-chart)
'red', "black", "green", "purple"
or the Hex notation, "coral"== "#FF7F50"
- R color cheatsheet :
<https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/colorPaletteCheatsheet.pdf> (w. examples)
- R palettes packages: grDevices and colorRamps, RcolorBrewer, colorspace;
- Interactive Tools: <http://colorbrewer2.org> ; use pal<- choose_palette ()



Sec 3.4: Plotting with Color

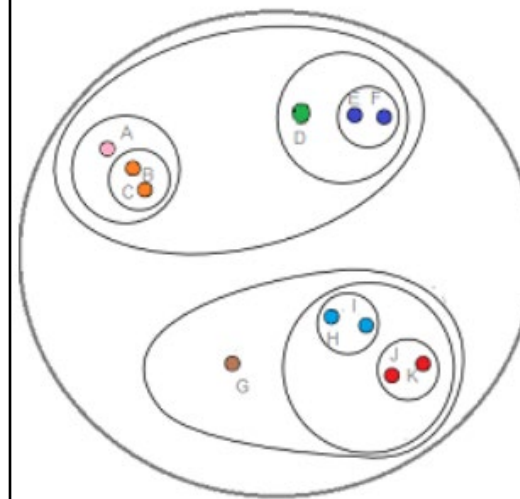
➤ Hierarchical Clustering / Dendrogram: Definition, Examples

Hierarchical clustering is where you build a cluster tree (a **dendrogram**) to represent data, where each group (or “node”) links to two or more successor groups. The groups are nested and organized as a tree, which ideally ends up as a meaningful classification scheme.

Each node in the cluster tree contains a group of similar data; Initially, each object is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters, continuing until there is just a single cluster. The tree gives a visual snapshot of the data contained in the whole set. The total number of clusters is not predetermined before you start the tree creation.

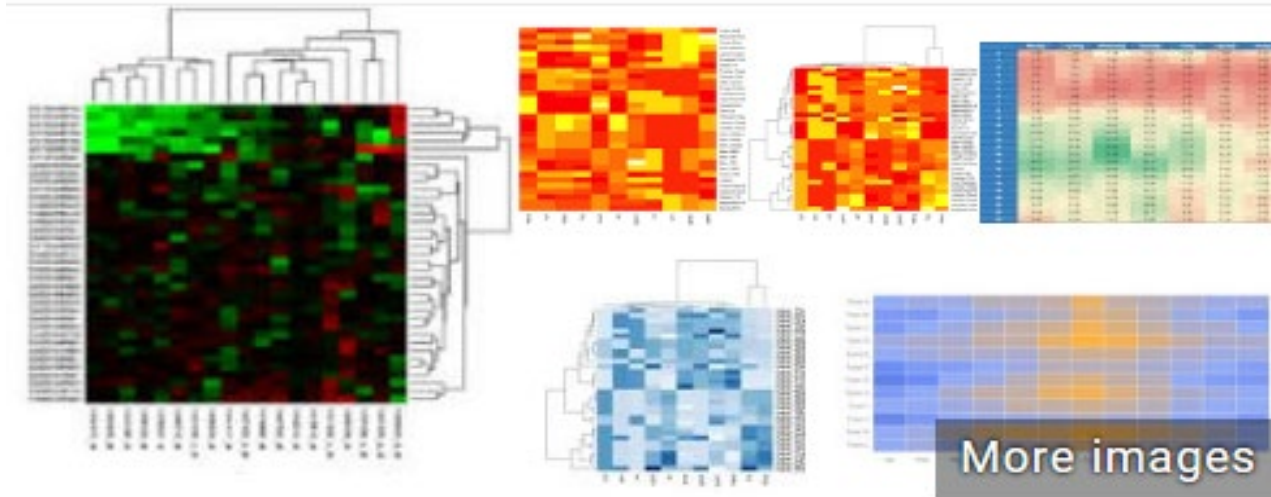
Unsupervised Learning.

<https://www.statisticshowto.datasciencecentral.com/hierarchical-clustering>



A dendrogram (right) representing nested clusters (left).

Sec C4.3.2 (p51) Heatmap (color-matrix)



Heat map



A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. "Heat map" is a newer term but shading matrices have existed for over a century.

[Wikipedia](#)

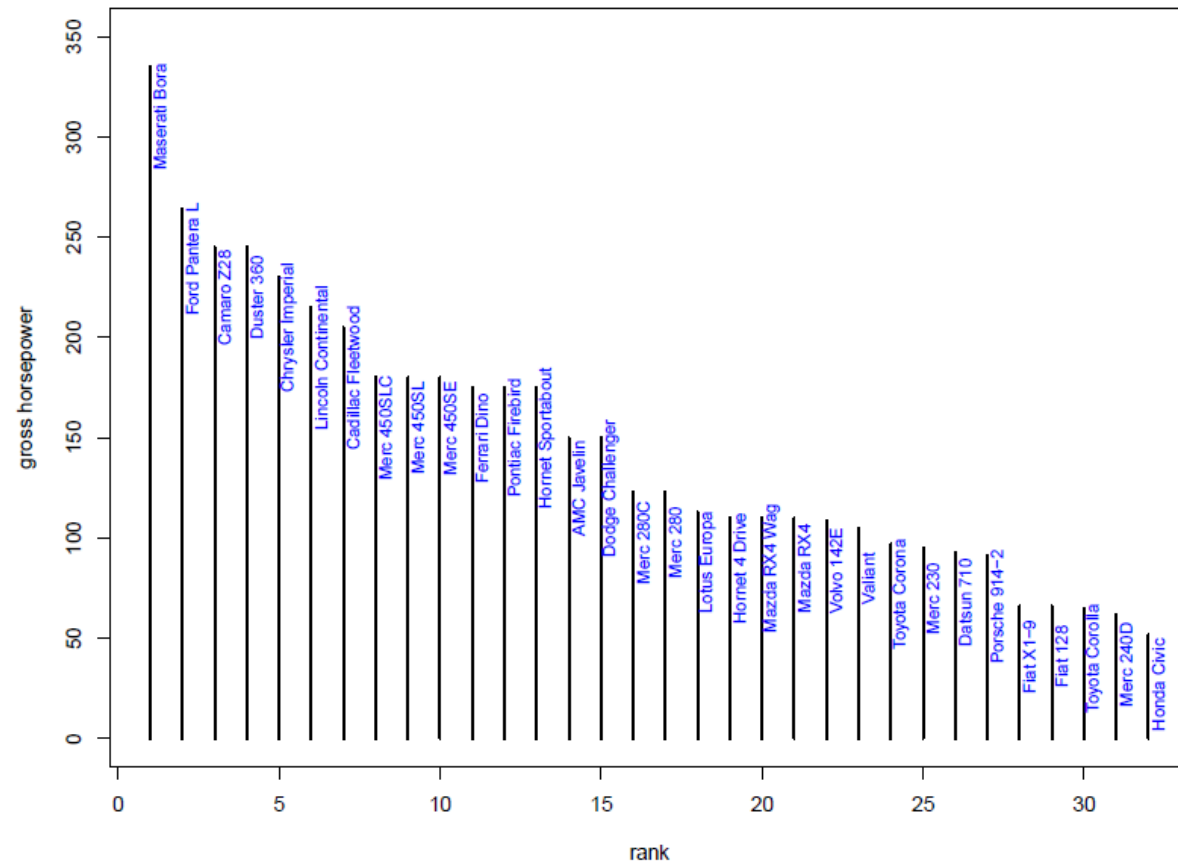


Sec D: R graphic Devices (static)

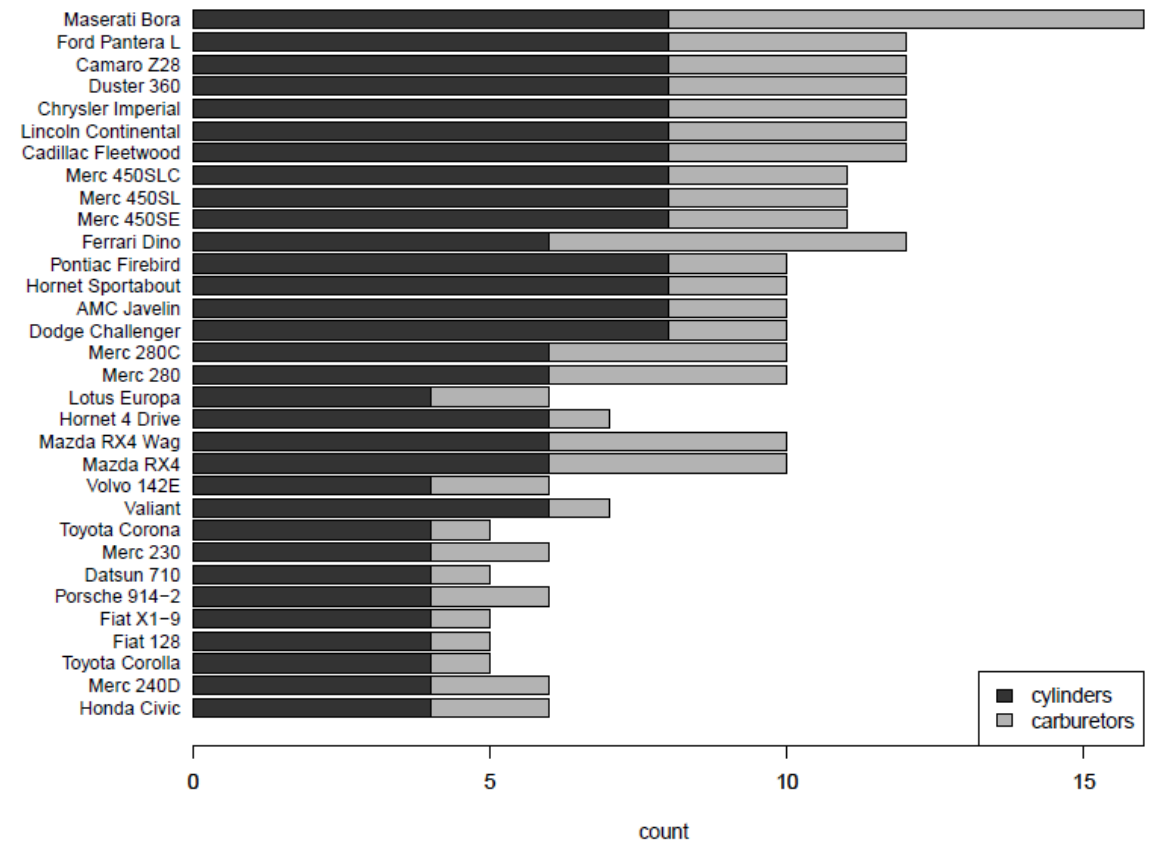
- Graphic output in **screen** device or **file** device: >? Devices
 - The following graphics devices are currently available:
 - [windows](#) The graphics device for Windows (on screen, to printer and to Windows metafile).
 - [pdf](#) Write PDF graphics commands to a file
 - [postscript](#) Writes PostScript graphics commands to a file
 - [xfig](#) Device for XFIG graphics file format
 - [bitmap](#) bitmap pseudo-device via Ghostscript (if available).
 - [pictex](#) Writes TeX/PicTeX graphics commands to a file (of historical interest only)
- The following devices will be functional if **R** was compiled to use them (they exist but will return with a warning on other systems):
- [cairo_pdf](#), [cairo_ps](#) PDF and PostScript devices based on cairo graphics.
 - [svg](#) SVG device based on cairo graphics.
 - [png](#) PNG bitmap device
 - [jpeg](#) JPEG bitmap device
 - [bmp](#) BMP bitmap device
 - [tiff](#) TIFF bitmap device

Ex3.3

Data Obtained in mid-1970s



Data ordered by gross horsepower



Sec E: Interactive Display

In R, the pipe operator is, as you have already seen, `%>%`: you can think of this operator as being similar to the `+` in a `ggplot2` statement. It takes the output of one statement and makes it the input of the next statement. When describing it, you can think of it as a "THEN".

```
iris %>%
```

```
  subset(Sepal.Length > 5) %>%
```

```
  aggregate(. ~ Species, ., mean)
```

```
# x %>% f(y) -> f(x, y)
```

➔ Pipeline: you take the Iris data, then you subset the data and then you aggregate the data

➔ `%>%` can be used in `ggvis` as well as data processing steps.

htmlwidgets for R

Interactivity: Any HTML format (document, notebook, presentation, or dashboard) can contain interactive components.

<https://www.htmlwidgets.org/> (see showcase)

Dygraphs: provides rich facilities for charting time-series data in R and includes support for many interactive features including series/point highlighting, zooming, etc.

Plotly: allows you to easily translate your ggplot2 graphics to an interactive web-based version:

https://images.plot.ly/plotly-documentation/images/r_cheat_sheet.pdf

Image can save as webpage- html file and include in R markdown file

Other interesting ones: Highcharter; d3heatmap, threejs, etc