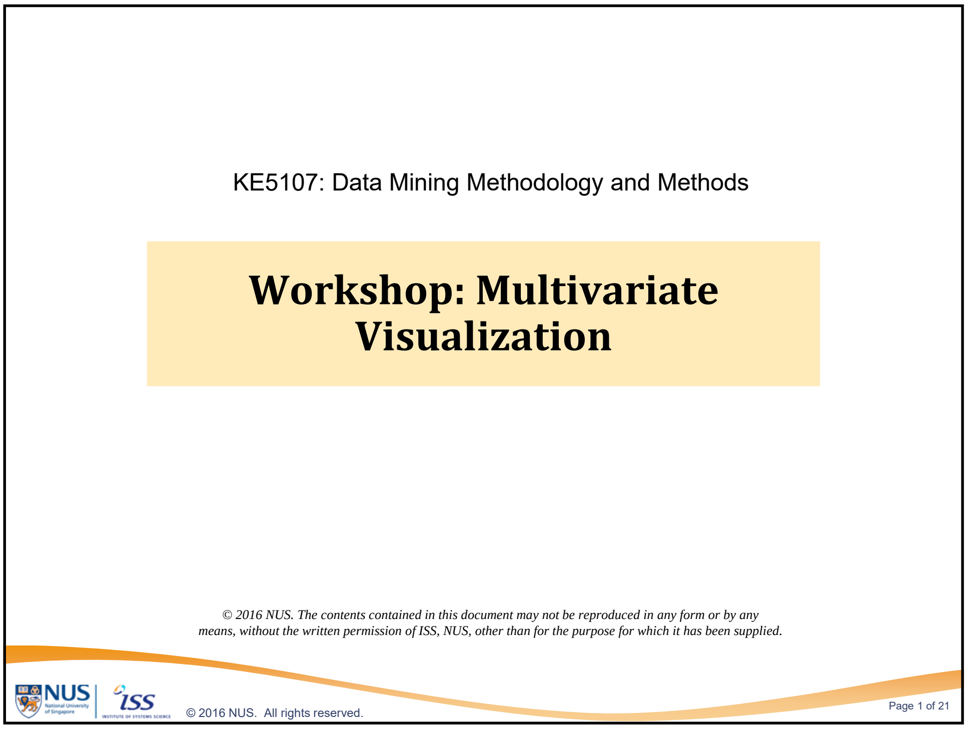
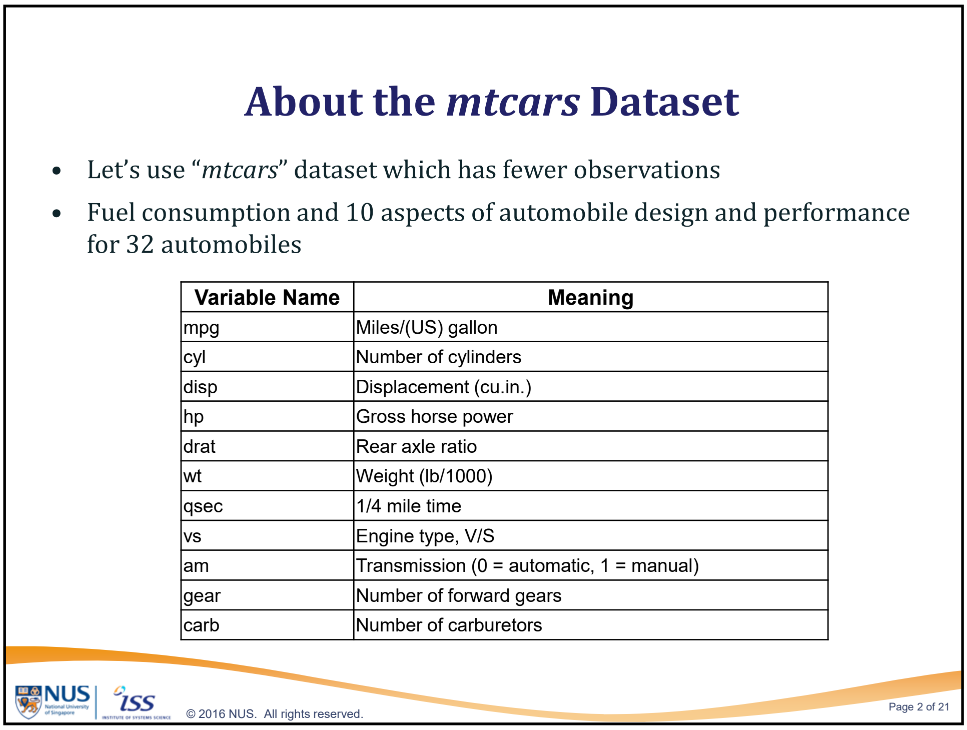


KE5107: Data Mining Methodology and Methods

Workshop: Multivariate Visualization

© 2016 NUS. The contents contained in this document may not be reproduced in any form or by any means, without the written permission of ISS, NUS, other than for the purpose for which it has been supplied.



About the *mtcars* Dataset

- Let’s use “*mtcars*” dataset which has fewer observations
- Fuel consumption and 10 aspects of automobile design and performance for 32 automobiles

Variable Name	Meaning
mpg	Miles/(US) gallon
cyl	Number of cylinders
disp	Displacement (cu.in.)
hp	Gross horse power
drat	Rear axle ratio
wt	Weight (lb/1000)
qsec	1/4 mile time
vs	Engine type, V/S
am	Transmission (0 = automatic, 1 = manual)
gear	Number of forward gears
carb	Number of carburetors

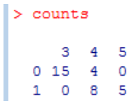
Overlay with Color: Stacked Bar Plot

- First get the counts

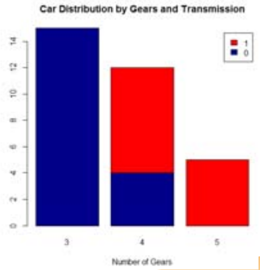
  counts <- table(mtcars$am, mtcars$gear)

- Then plot

  barplot(counts, main="Car Distribution by Gears and Transmission",
  xlab="Number of Gears", col=c("darkblue","red"),
  legend.text = rownames(counts), names.arg=colnames(counts))



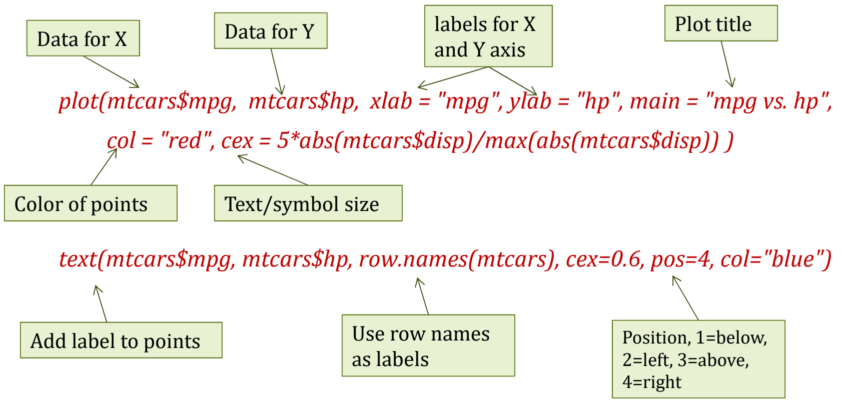
```
> counts

     3  4  5
  0 15  4  0
  1  0  8  5
```



© 2016 NUS. All rights reserved.

Overlay with Size: Bubble Plot

- Add a third dimension "disp" to the size of the points in a scatter plot



plot(mtcars$mpg, mtcars$hp, xlab = "mpg", ylab = "hp", main = "mpg vs. hp",
*col = "red", cex = 5*abs(mtcars$disp)/max(abs(mtcars$disp)))*

Data for X

Data for Y

labels for X and Y axis

Plot title

Color of points

Text/symbol size

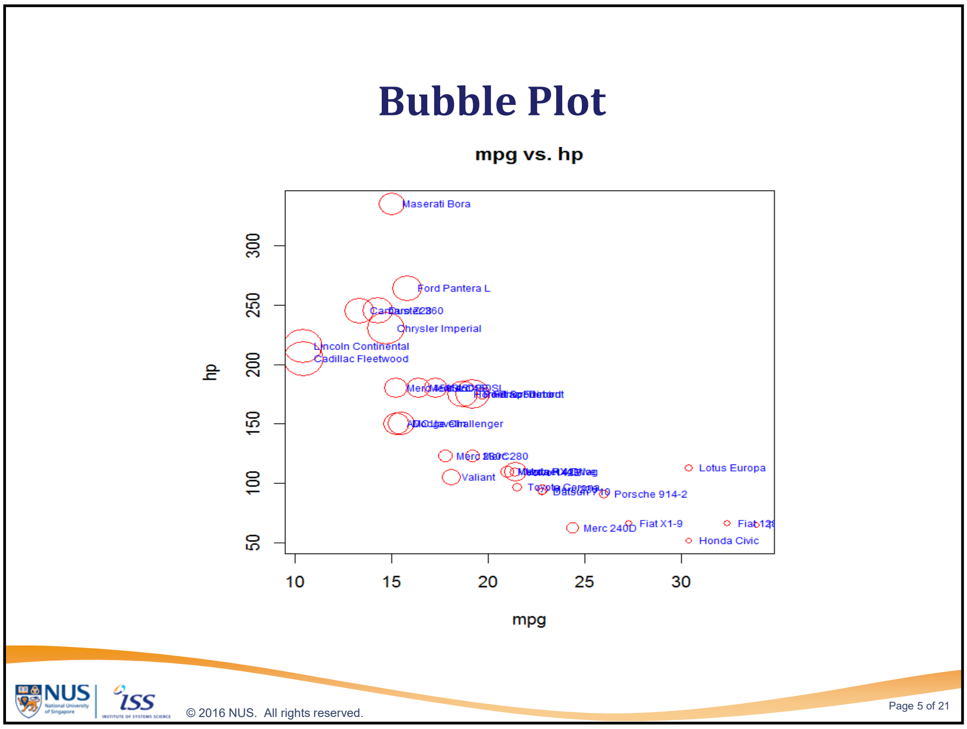
text(mtcars$mpg, mtcars$hp, row.names(mtcars), cex=0.6, pos=4, col="blue")

Add label to points

Use row names as labels

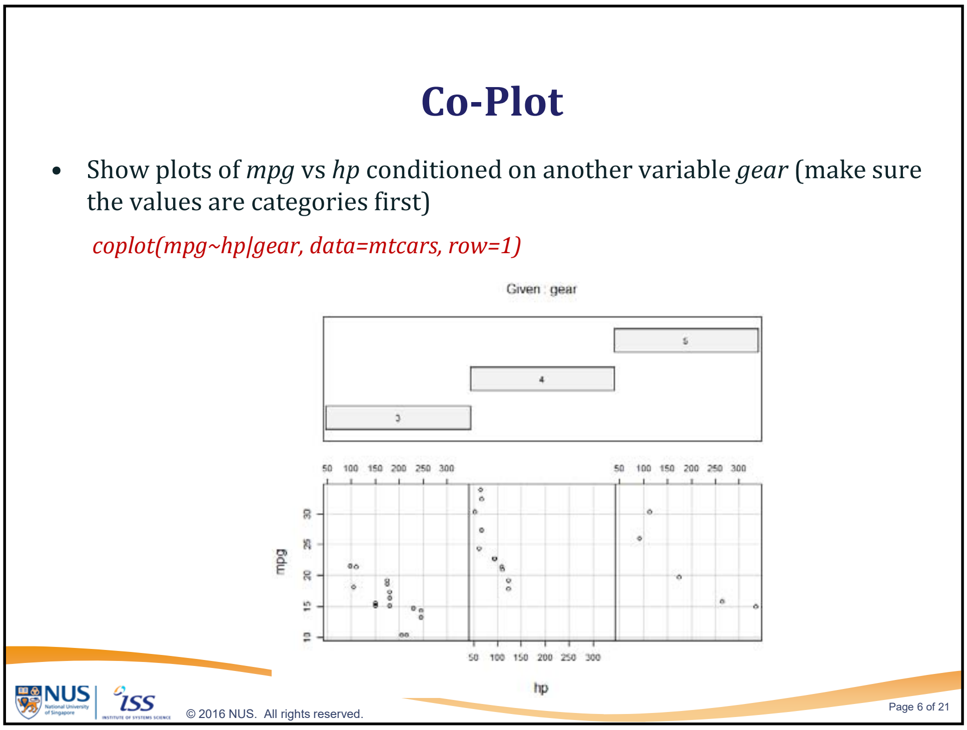
Position, 1=below, 2=left, 3=above, 4=right

© 2016 NUS. All rights reserved.



Bubble Plot

© 2016 NUS. All rights reserved.

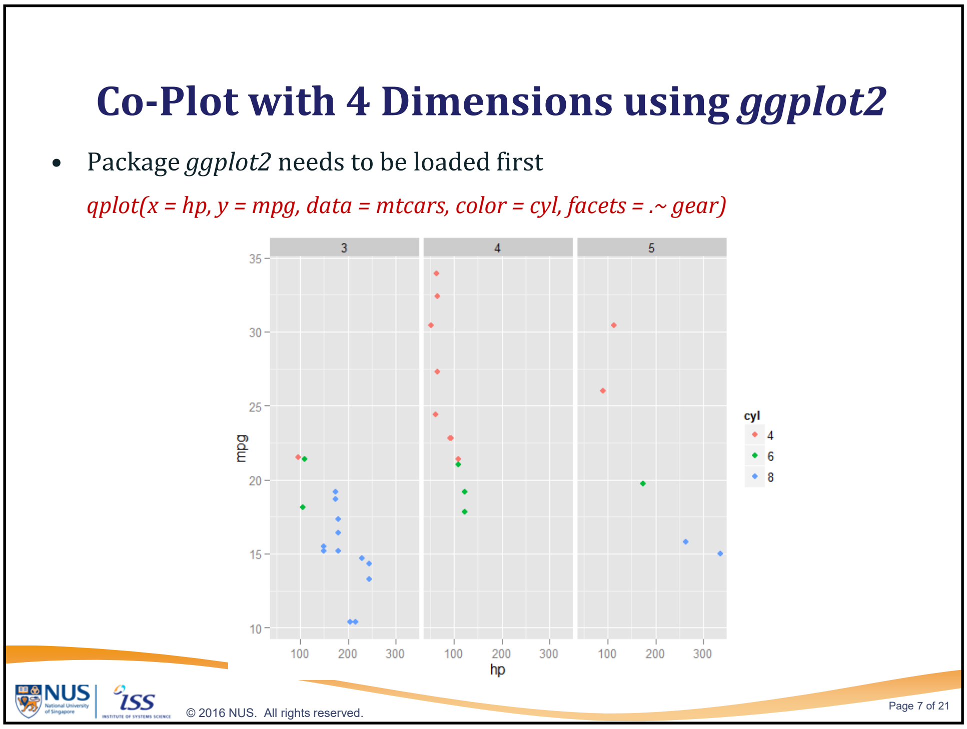


Co-Plot

- Show plots of *mpg* vs *hp* conditioned on another variable *gear* (make sure the values are categories first)

coplot(mpg~hp|gear, data=mtcars, row=1)

© 2016 NUS. All rights reserved.

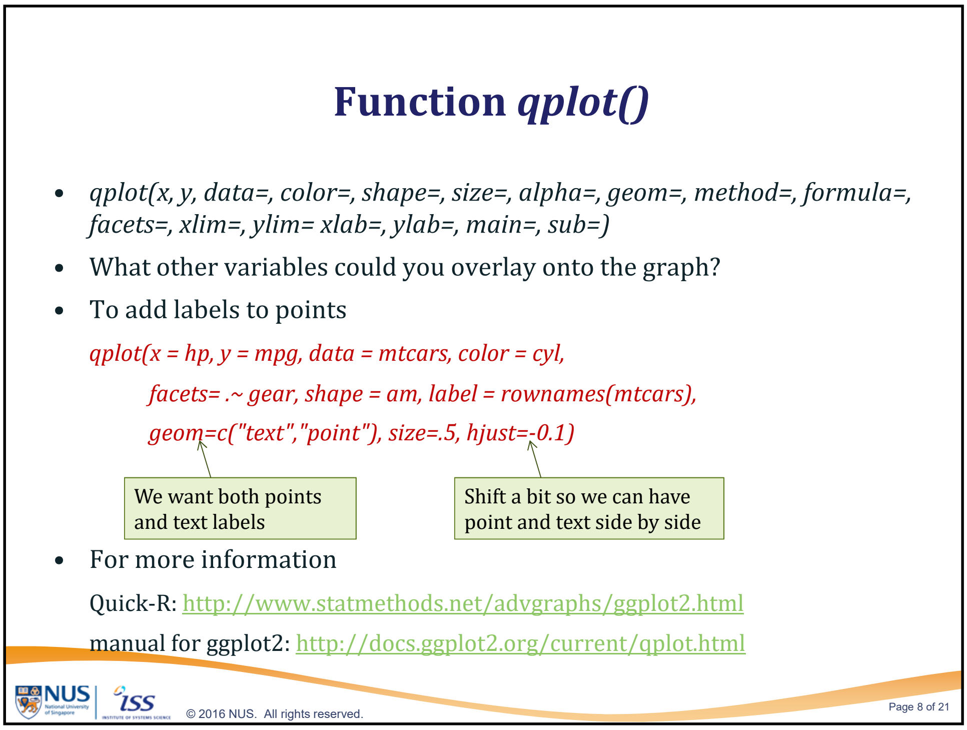


Co-Plot with 4 Dimensions using *ggplot2*

- Package *ggplot2* needs to be loaded first

  qplot(x = hp, y = mpg, data = mtcars, color = cyl, facets = .~ gear)

© 2016 NUS. All rights reserved.



Function *qplot()*

- *qplot(x, y, data=, color=, shape=, size=, alpha=, geom=, method=, formula=, facets=, xlim=, ylim= xlab=, ylab=, main=, sub=)*
- What other variables could you overlay onto the graph?
- To add labels to points

  qplot(x = hp, y = mpg, data = mtcars, color = cyl,

  facets= .~ gear, shape = am, label = rownames(mtcars),

  geom=c("text","point"), size=.5, hjust=-0.1)

  We want both points and text labels

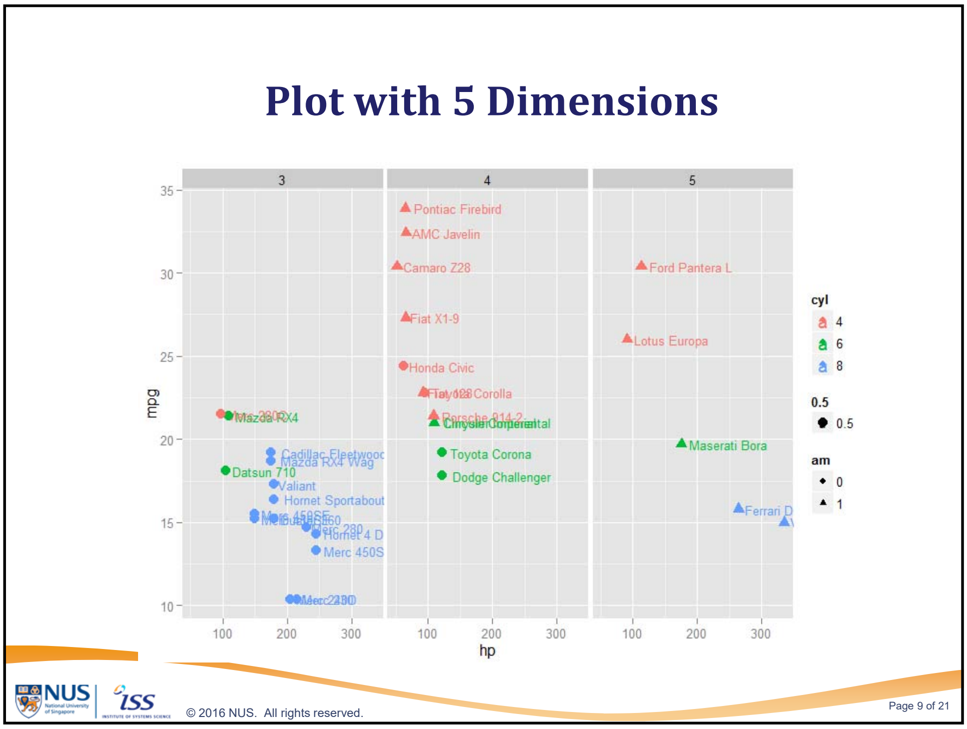
  Shift a bit so we can have point and text side by side

- For more information

  Quick-R: http://www.statmethods.net/advgraphs/ggplot2.html

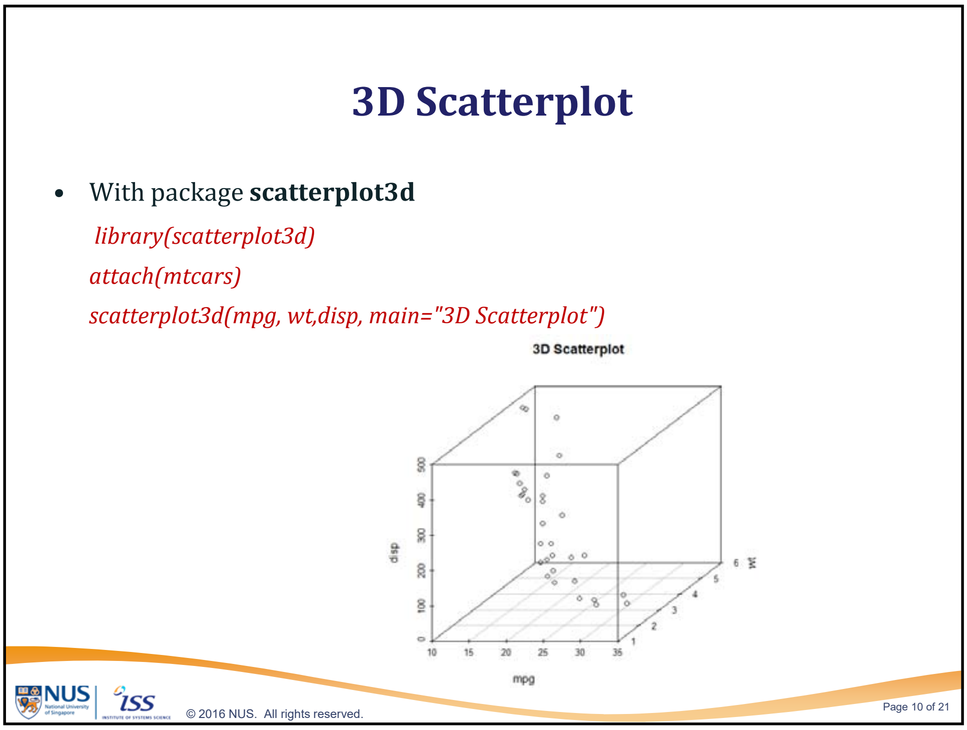
  manual for ggplot2: http://docs.ggplot2.org/current/qplot.html

© 2016 NUS. All rights reserved.



Plot with 5 Dimensions

© 2016 NUS. All rights reserved.



3D Scatterplot

- With package **scatterplot3d**

library(scatterplot3d)

attach(mtcars)

scatterplot3d(mpg, wt,disp, main="3D Scatterplot")

© 2016 NUS. All rights reserved.

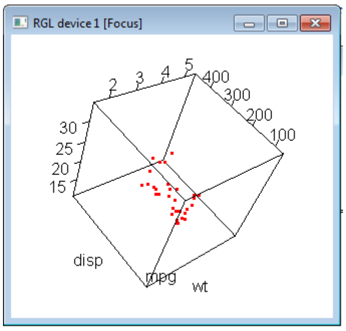
Interactive 3D Scatterplot

- With package **rgl**

  library(rgl)

  plot3d(wt, disp, mpg, col="red", size=3)

- Use mouse to rotate the graph



© 2016 NUS. All rights reserved.

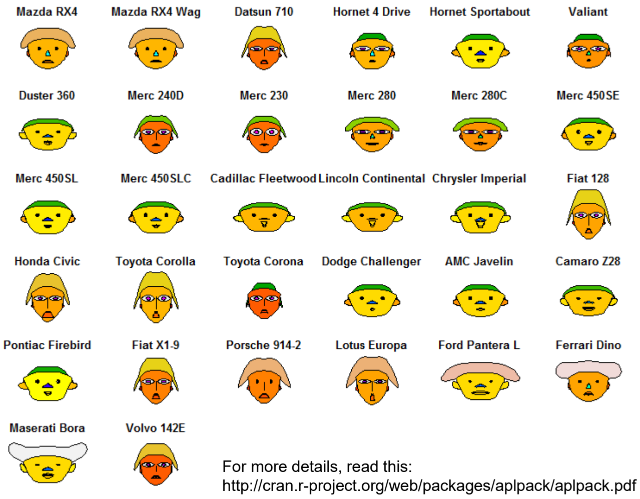
Chernoff Faces

install.packages("aplpack")

library(aplpack)

faces(mtcars)

```
modified item       Var
"height of face   " "mpg"
"width of face    " "cyl"
"structure of face" "disp"
"height of mouth  " "hp"
"width of mouth   " "drat"
"smiling          " "wt"
"height of eyes   " "qsec"
"width of eyes    " "vs"
"height of hair   " "am"
"width of hair   "  "gear"
"style of hair   "  "carb"
"height of nose  "  "mpg"
"width of nose   "  "cyl"
"width of ear    "  "disp"
"height of ear   "  "hp"
```



For more details, read this:
http://cran.r-project.org/web/packages/aplpack/aplpack.pdf

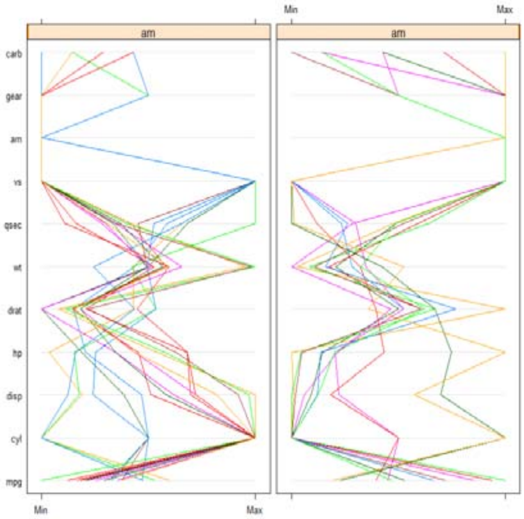
© 2016 NUS. All rights reserved.

Parallel Coordinates Plot

- With package **lattice**

  library(lattice)

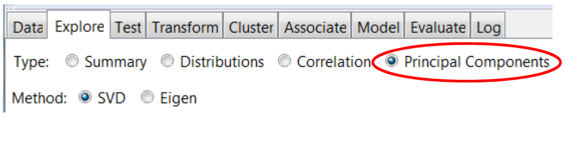
  parallelplot(~mtcars | am, mtcars)



© 2016 NUS. All rights reserved.

Principal Component Analysis

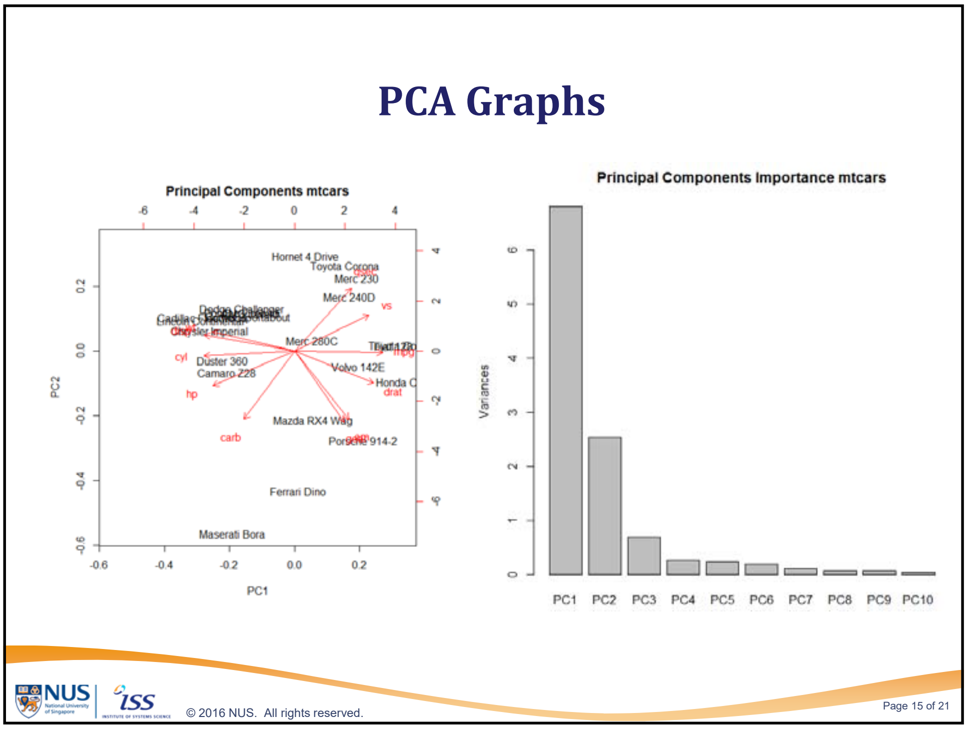
- Let's use *rattle* to do this. Load *mtcars*. Use all variables as numeric and input
- Select "Principal Components" from *Explore* Tab for PCA of numerical variables
- Two methods:
  - SVD – prcomp()
  - Eigen – princomp()



Type: ○ Summary ○ Distributions ○ Correlation ● Principal Components

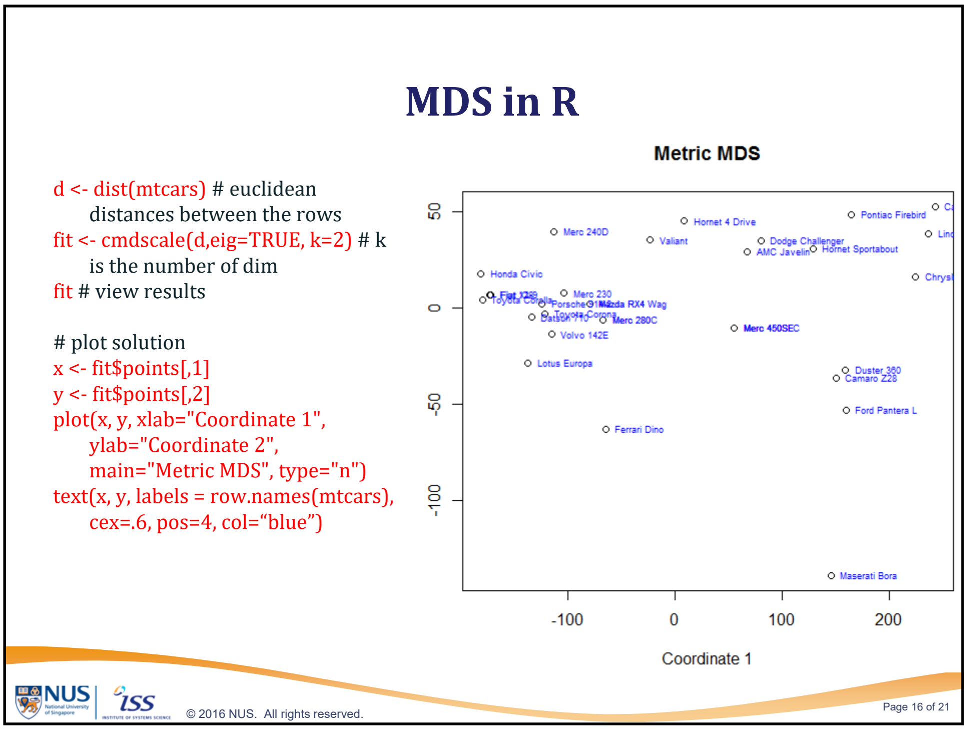
Method: ● SVD ○ Eigen

© 2016 NUS. All rights reserved.



PCA Graphs

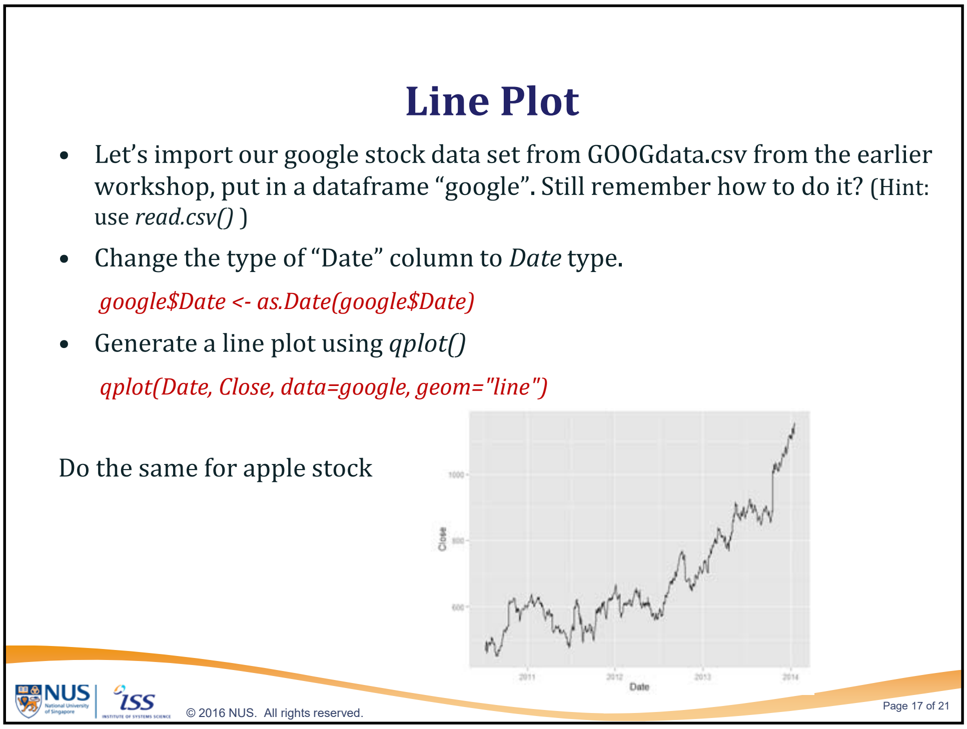
© 2016 NUS. All rights reserved.



MDS in R

```
d <- dist(mtcars) # euclidean
    distances between the rows
fit <- cmdscale(d,eig=TRUE, k=2) # k
    is the number of dim
fit # view results

# plot solution
x <- fit$points[,1]
y <- fit$points[,2]
plot(x, y, xlab="Coordinate 1",
    ylab="Coordinate 2",
    main="Metric MDS", type="n")
text(x, y, labels = row.names(mtcars),
    cex=.6, pos=4, col="blue")
```

© 2016 NUS. All rights reserved.



Line Plot

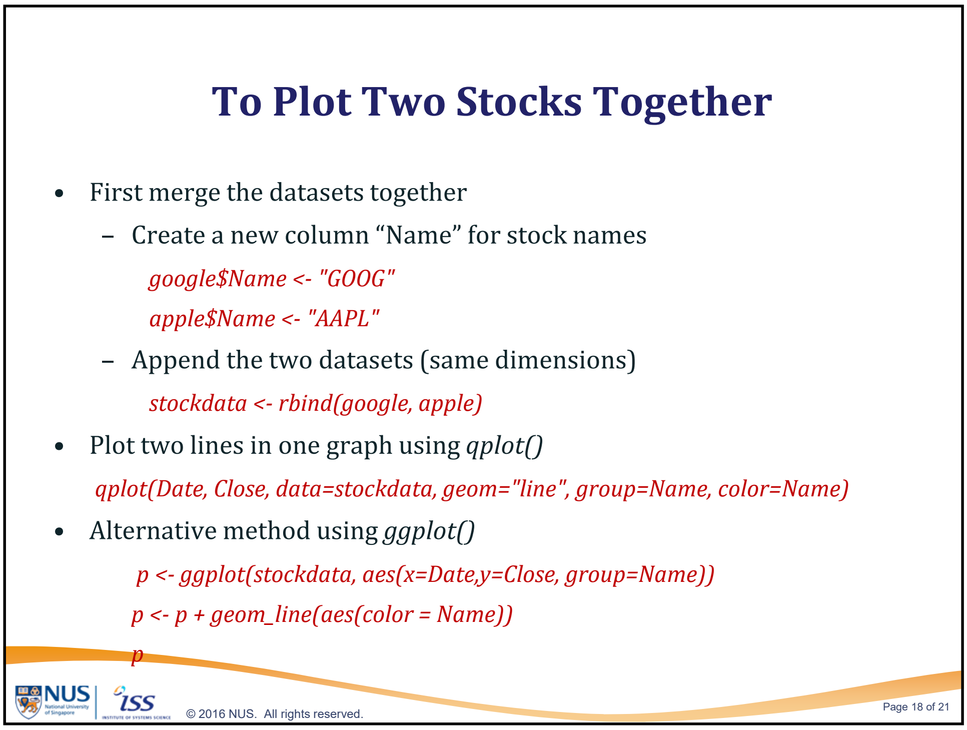
- Let's import our google stock data set from GOOGdata.csv from the earlier workshop, put in a dataframe "google". Still remember how to do it? (Hint: use *read.csv()*)
- Change the type of "Date" column to *Date* type.

  google$Date <- as.Date(google$Date)
- Generate a line plot using *qplot()*

  qplot(Date, Close, data=google, geom="line")

Do the same for apple stock

© 2016 NUS. All rights reserved.



To Plot Two Stocks Together

- First merge the datasets together
  - Create a new column "Name" for stock names

    google$Name <- "GOOG"

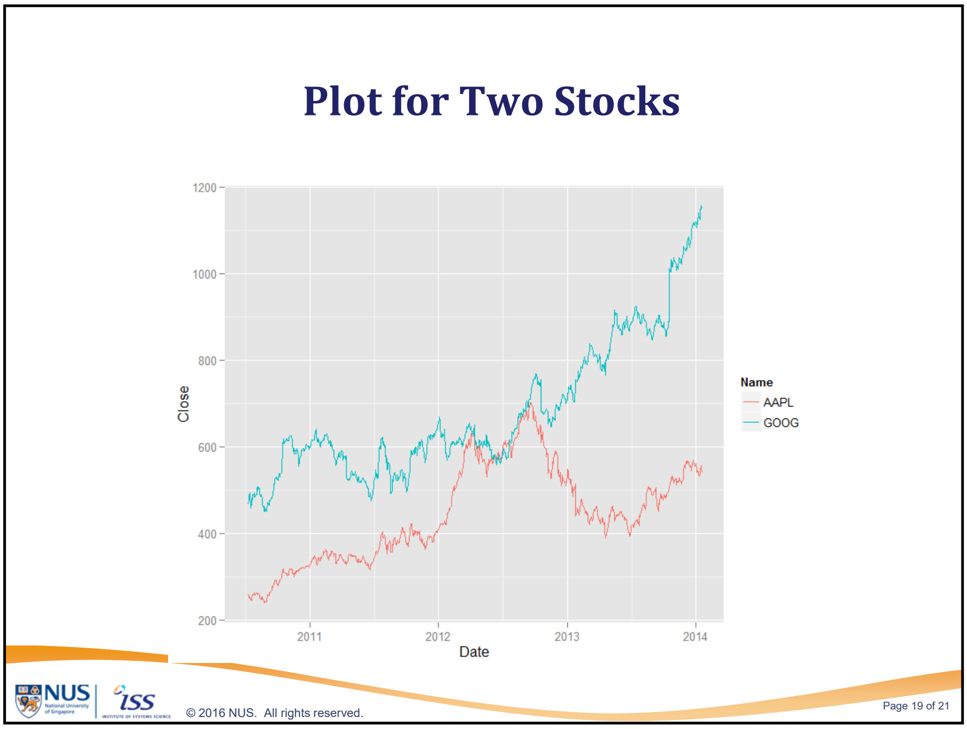
    apple$Name <- "AAPL"
  - Append the two datasets (same dimensions)

    stockdata <- rbind(google, apple)
- Plot two lines in one graph using *qplot()*

  qplot(Date, Close, data=stockdata, geom="line", group=Name, color=Name)
- Alternative method using *ggplot()*

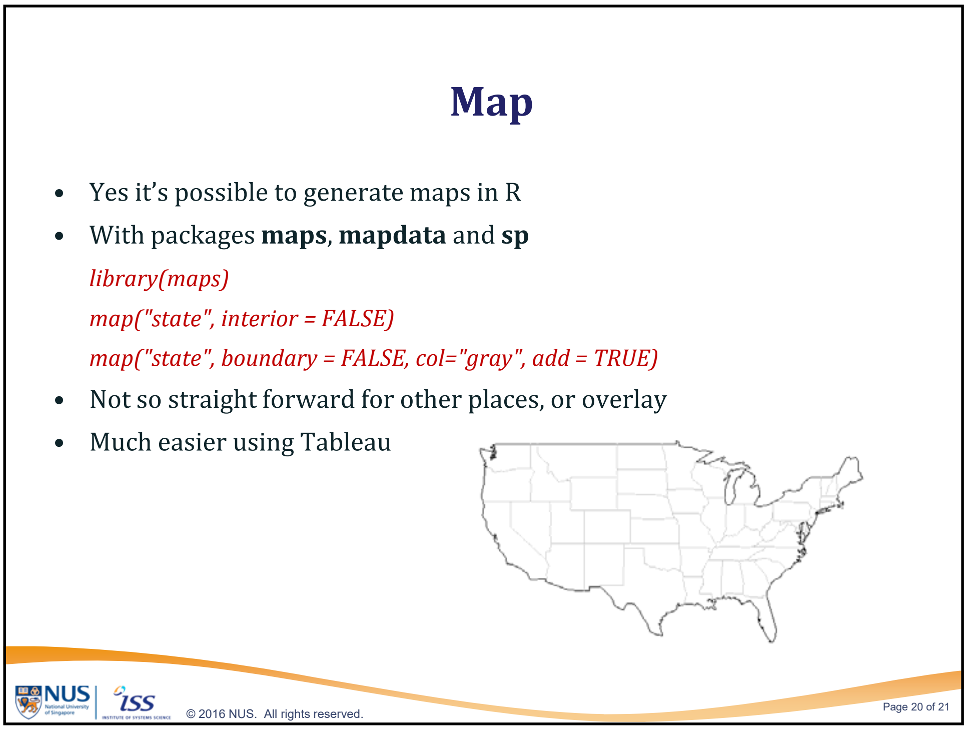
```
 p <- ggplot(stockdata, aes(x=Date,y=Close, group=Name))
p <- p + geom_line(aes(color = Name))
p
```

© 2016 NUS. All rights reserved.



Plot for Two Stocks

© 2016 NUS. All rights reserved.



Map

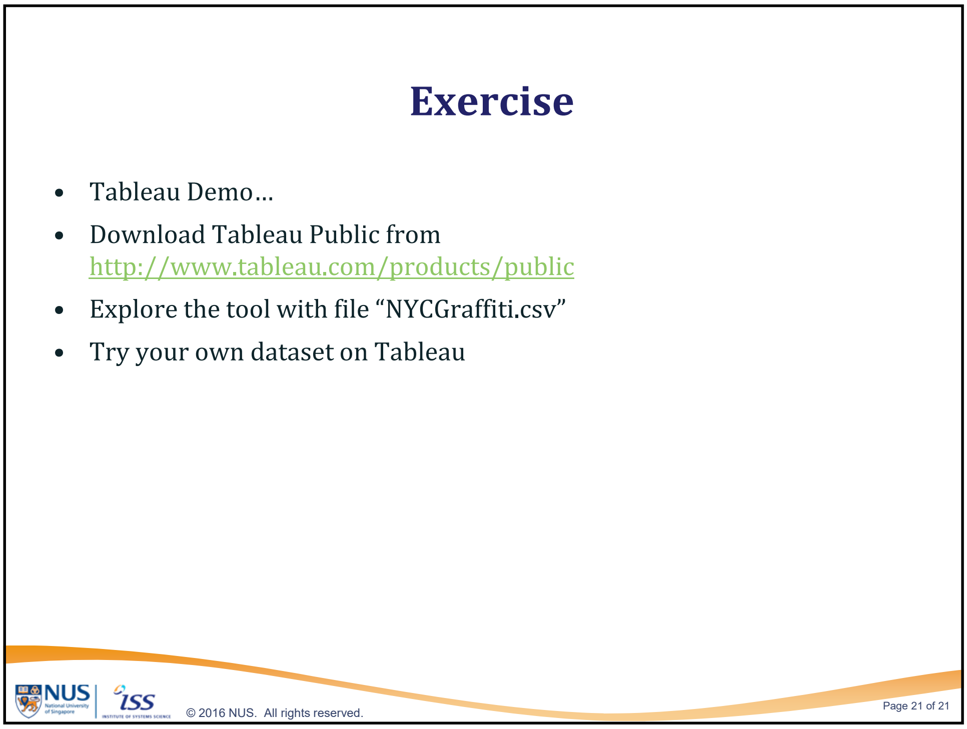
- Yes it's possible to generate maps in R
- With packages **maps**, **mapdata** and **sp**

  library(maps)

  map("state", interior = FALSE)

  map("state", boundary = FALSE, col="gray", add = TRUE)
- Not so straight forward for other places, or overlay
- Much easier using Tableau

© 2016 NUS. All rights reserved.



Exercise

- Tableau Demo…
- Download Tableau Public from http://www.tableau.com/products/public
- Explore the tool with file “NYCGraffiti.csv”
- Try your own dataset on Tableau

© 2016 NUS. All rights reserved.