Welcome to the Plant Breeding and Genomics Webinar Series

Today's Presenter: Dr. Heather Merk

Presentation & Supplemental Files: http://www.extension.org/pages/60427

Co-Hosts: John McQueen & Heather Merk

PBG home page:
www.eXtension.org/plant_breeding_genomics

Sign up for PBG News: http://pbgworks.org

Please fill out the survey evaluation! (You will be contacted via email)

Watch past webinars and sign up for future webinars!

http://www.extension.org/pages/60426

Introduction to R Statistical Software: Application to Plant Breeding

Presenter: Dr. Heather L. Merk

merk.9@osu.edu

The Ohio State University, OARDC

Overview

- Why R?
- Where to Obtain R
- How to Perform Basic Commands
- Sample Analyses
- How to Obtain Help
- How to Learn More

Learning Objectives

At the end of this webinar you should be able to do the following using R…

- Install and run R. Find R packages, install, and load them.
- Read in data and visualize distribution
- Test if there are differences between varieties (ANOVA using linear regression)
- Distinguish varieties (Means and T-test)
- Estimate variance components
- Use loops to simplify analysis

R Overview

- Open-source programming language for statistical analysis and graphing
- Based on S (developed by Bell Labs, the developers of Unix. You will see similarities)
- Provides language, tool, and environment in one
- Functions and analysis stored as objects, allowing for function modification and model building
- Many packages for specific applications are already available

Why R?

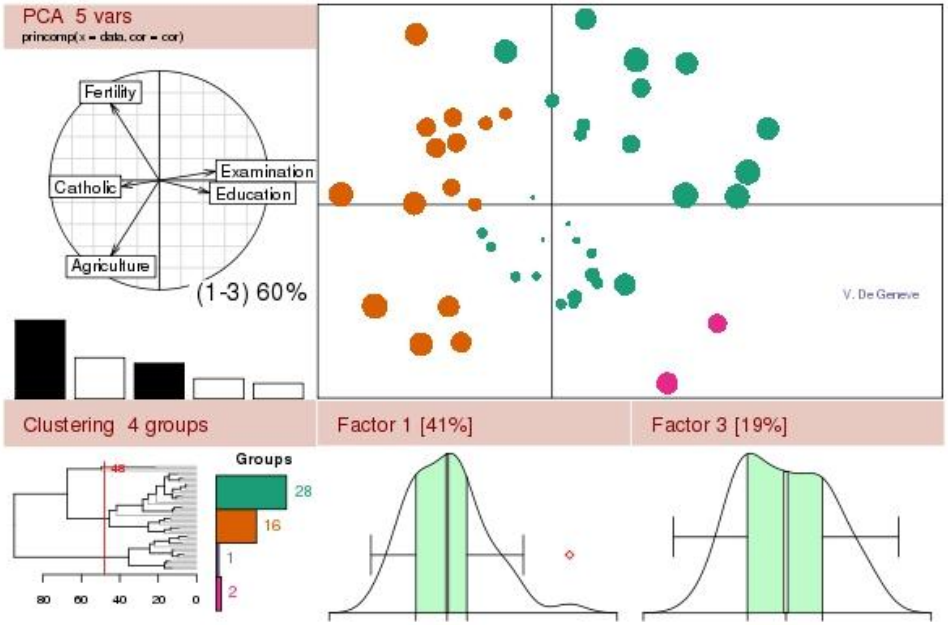
www.r-project.org/foundation

- R is free!
- Powerful software
- Publication quality figures
- Built-in help
- Many resources

Obtain R

www.r-project.org

The R Project for Statistical Computing



Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.
- If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

Choose a CRAN Mirror

- CRAN = Comprehensive R Archive Network
- Select the mirror site closest to you

CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found here: main page, windows release, windows old release.

Argentina

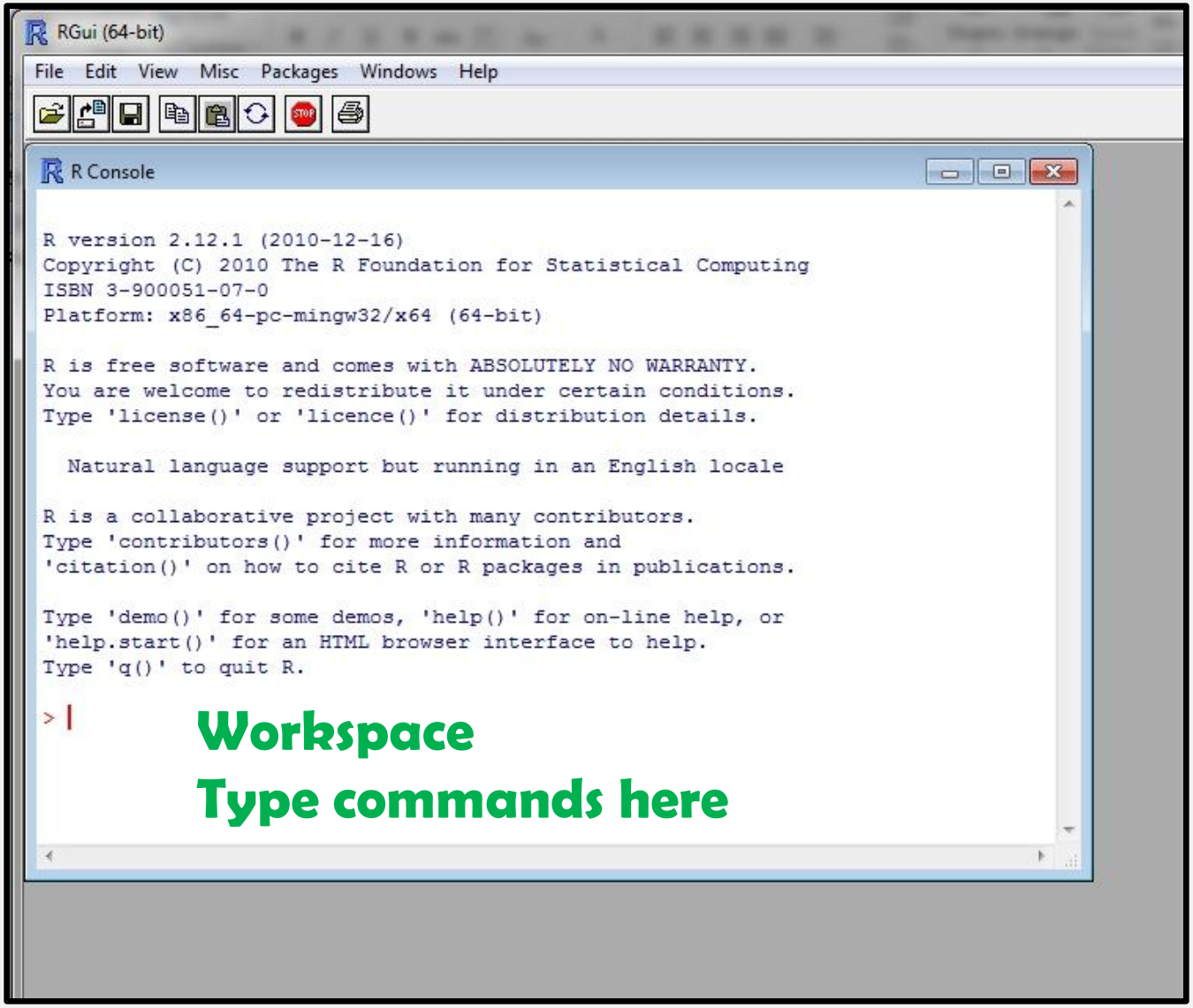
http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
http://r.mirror.mendoza-conicet.gob.ar/	CONICET Mendoza

Australia

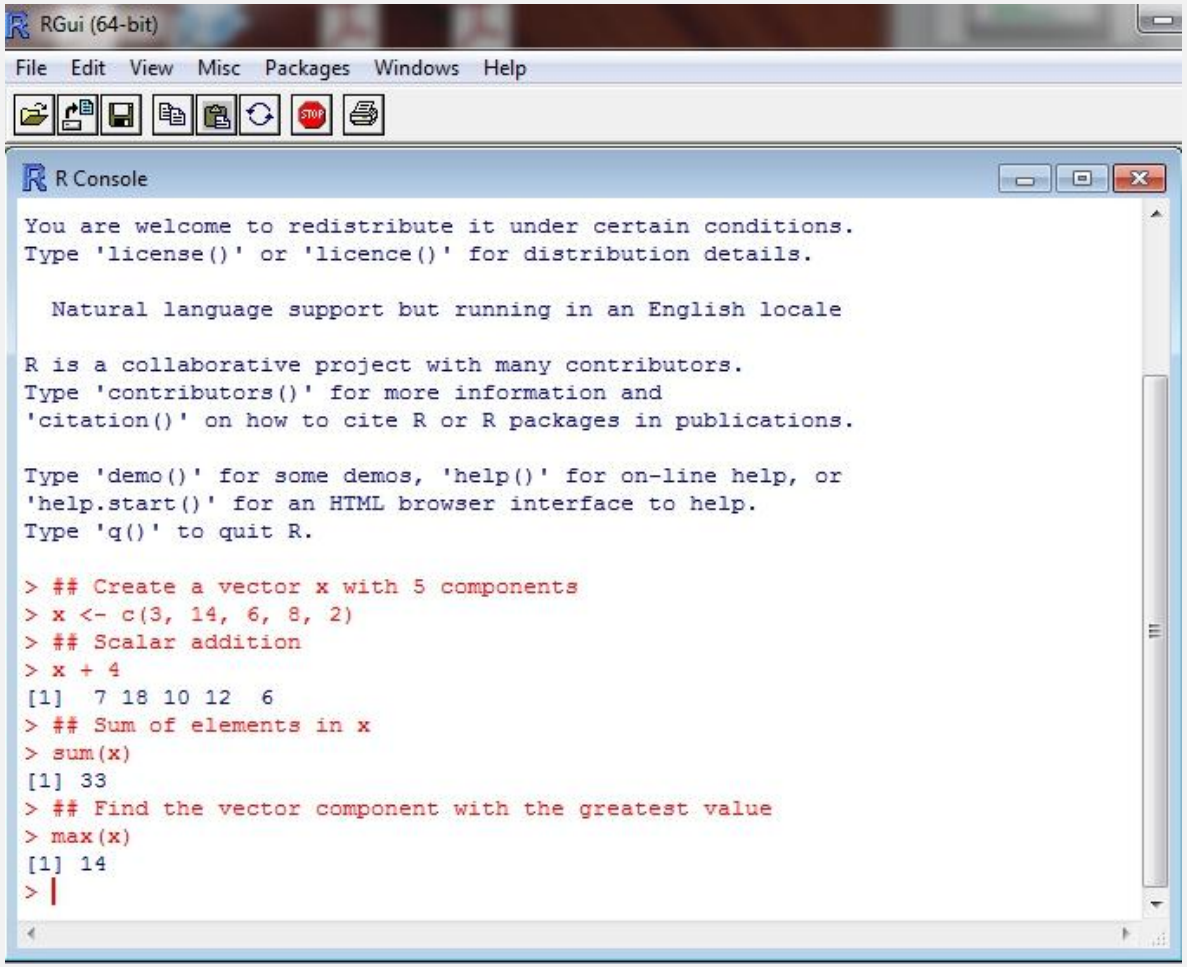
http://cran.csiro.au/	CSIRO
http://cran.ms.unimelb.edu.au/	University of Melbourne

Now That You Have R, the Fun Begins!

Command line Interface



Basic Commands in R



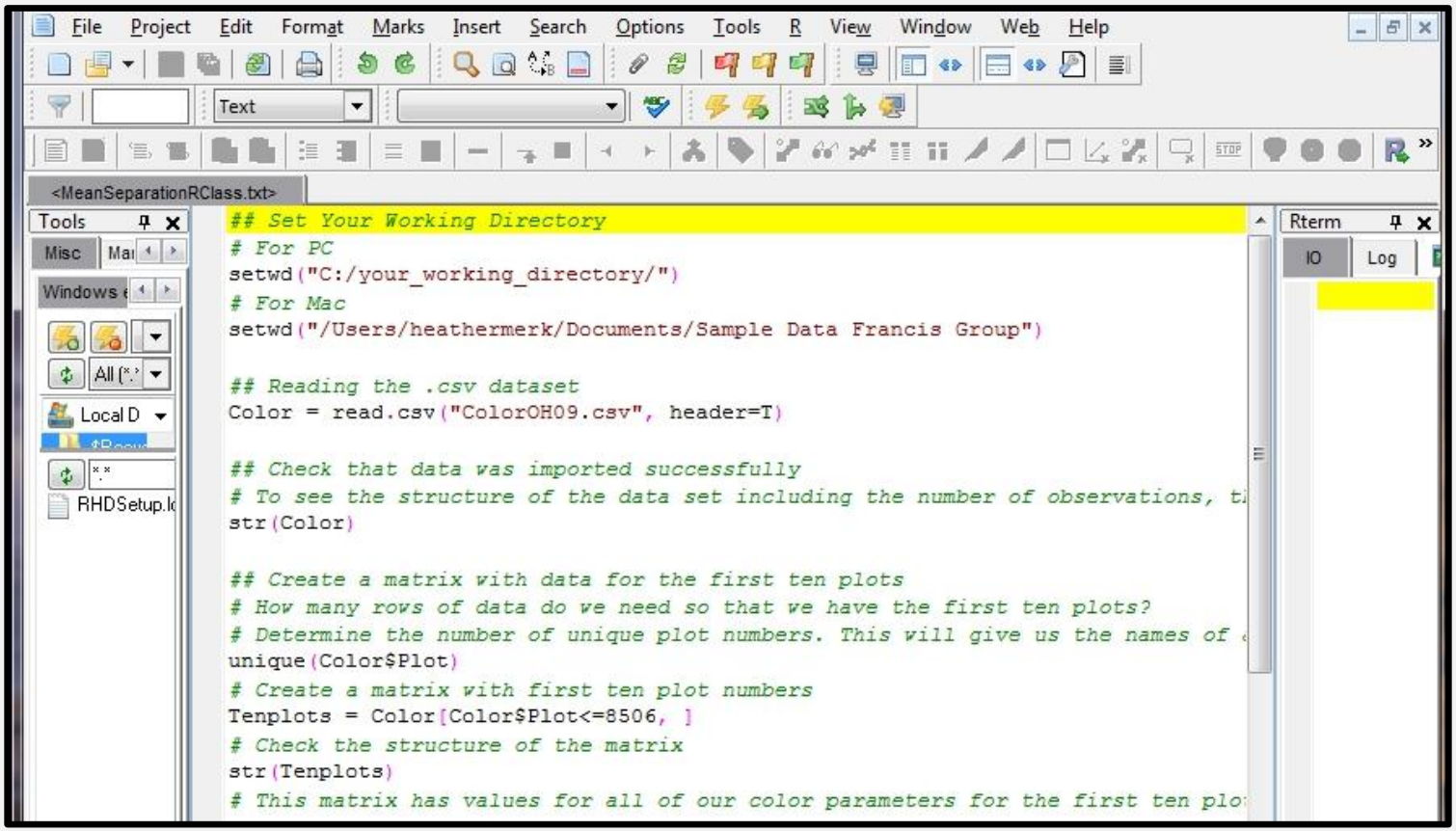
- R is case sensitive
- # comment follows
- <- or = assignment operator
- c concatenate

Practice Entering Commands!

- Kim, D.Y. R basics [Online]. Illinois State University. Available at: http://math.illinoisstate.edu/dhkim/rstuff/rtutor.html (verified 8 Sept 2011).

Text Editors

- Alternative to typing in command line
- Write scripts that can easily be saved and recalled
- Mac – built-in color text editor
- PC – Tinn-R (Tinn is not Notepad), difficulties with Windows 7 and with newer versions of R, http://sciviews.org/Tinn-R

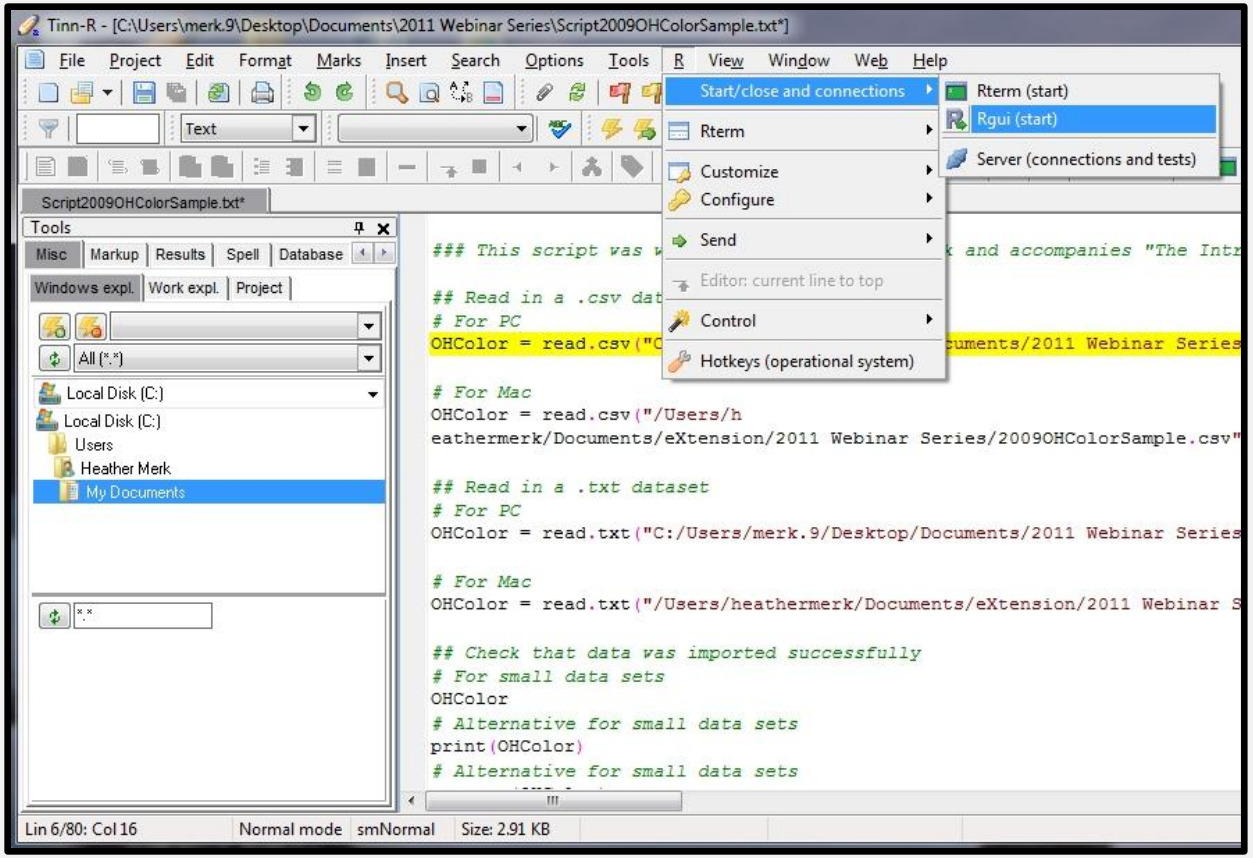


Running a Script on a PC

- Copy and paste into R console

OR

- Tinn-R – Open R, click on a line of your script, press Ctrl + Enter



Running a Script on a Mac

- Copy and paste into R console

OR

- Highlight line, press Command + Enter

Script2009OHColorSample.txt

<functions> Help search

```
## Read in a .csv dataset
# For PC
OHColor = read.csv("C:/Users/merk.9/Desktop/Documents/2011 Webinar Series/2009OHColorSample.csv", header=T)

# For Mac
OHColor = read.csv("/Users/heathermerk/Documents/eXtension/2011 Webinar Series/2009OHColorSample.csv",
 header=T)

## Read in a .txt dataset
# For PC
OHColor = read.txt("C:/Users/merk.9/Desktop/Documents/2011 Webinar Series/2009OHColorSample.txt", header=T)

# For Mac
OHColor = read.txt("/Users/heathermerk/Documents/eXtension/2011 Webinar Series/2009OHColorSample.txt",
 header=T)

## Check that data was imported successfully
# For small data sets
OHColor
# Alternative for small data sets
print(OHColor)
# Alternative for small data sets
summary(OHColor)

# For large data sets
# To see the first rows of data including header
head(OHColor)
# To see the last rows of data including tail
tail(OHColor)
# To see the structure of the data set including the number of observations, the number of variables, the
 variable names, the number of levels of each variable
str(OHColor)
```

read.csv(file, header = TRUE, sep = ",", quote = "\"", dec = ".", fill = TRUE, comment.char = "", ...)

Sample Data

- SolCAP Phenotypic data
- Processing tomato fruit shape, color, quality data
- Scanned images analyzed with Tomato Analyzer software
- 2009OHColorSample.xls has color data from one year in one location
- 2010OHColorSample.xls has color data from one year in one location

Data format

- Tab delimited or csv files
- Flat file
- Missing data NA
- Headers – no spaces, no commas (csv), begin with a letter

	A	B	C	D	E	F	G	H	I	J	K	L
1	Line	Rep	Year	Loc	Param1	Param2	Avgred	Avggreen	Avgblue	Avglum	Avgl	Avga
2	SCT_0001	1	2009	OH	0.81	94.87	142.86	64.24	48.14	89.87	37.34	29.28
3	SCT_0001	1	2009	OH	1.55	91.92	132.52	60.50	46.36	84.14	34.86	27.14
4	SCT_0001	1	2009	OH	1.17	87.32	142.92	68.52	54.04	92.62	38.35	27.46
5	SCT_0001	1	2009	OH	1.49	94.71	125.71	56.97	44.78	80.18	32.98	26.26
6	SCT_0001	1	2009	OH	1.26	90.78	131.81	61.14	47.31	84.20	34.91	26.60
7	SCT_0001	1	2009	OH	0.98	92.99	141.71	64.88	51.72	90.97	37.33	28.75
8	SCT_0001	1	2009	OH	0.47	93.45	139.05	62.67	47.01	87.53	36.38	28.58
9	SCT_0001	1	2009	OH	0.62	95.46	127.42	57.03	44.96	81.07	33.25	26.91
10	SCT_0001	1	2009	OH	1.00	93.42	140.31	63.09	48.91	89.02	36.70	28.94
11	SCT_0002	1	2009	OH	0.99	88.47	132.11	64.34	50.16	85.70	35.69	25.22

Importing Data

- Script – Script2009OHColorSample.txt
- read.csv("filename.csv", header=T)
- read.txt("filename.txt", header=T)
- For PC – note the direction of the slashes

Script2009OHColorSample.txt

<functions>

Help search

```
## Read in a .csv dataset
# For PC
OHColor = read.csv("C:/Users/merk.9/Desktop/Documents/2011 Webinar Series/2009OHColorSample.csv", header=T)

# For Mac
OHColor = read.csv("/Users/heathermerk/Documents/eXtension/2011 Webinar Series/2009OHColorSample.csv",
 header=T)

## Read in a .txt dataset
# For PC
OHColor = read.txt("C:/Users/merk.9/Desktop/Documents/2011 Webinar Series/2009OHColorSample.txt", header=T)

# For Mac
OHColor = read.txt("/Users/heathermerk/Documents/eXtension/2011 Webinar Series/2009OHColorSample.txt",
 header=T)
```

Learn More – Importing Data

- R Development Core Team. R Data import/export [Online]. The Comprehensive R Archive Network. Available at: http://cran.r-project.org (verified 9 Sept 2011).

Checking Data – Small Data Sets ONLY!

- object or print(object)
- summary(object)

```
## Check that data was imported successfully
# For small data sets
OHColor
# Alternative for small data sets
print(OHColor)
# Alternative for small data sets
summary(OHColor)
```

Checking Data – Head & Tail Commands

- head(object)
  - See first rows of data including header
- tail(object)
  - See last rows of data including header

```
> OHColor = read.csv("C:/Users/merk.9/Desktop/Documents/2011 Webinar Series/2009OHColorSample.csv", header=T)
> # To see the first rows of data including header
> head(OHColor)
      Line Rep Year Loc Param1 Param2 Avgred Avggreen Avgblue Avglum  Avgl
1 SCT_0001   1 2009  OH   0.81  94.87 142.86    64.24   48.14  89.87 37.34
2 SCT_0001   1 2009  OH   1.55  91.92 132.52    60.50   46.36  84.14 34.86
3 SCT_0001   1 2009  OH   1.17  87.32 142.92    68.52   54.04  92.62 38.35
4 SCT_0001   1 2009  OH   1.49  94.71 125.71    56.97   44.78  80.18 32.98
5 SCT_0001   1 2009  OH   1.26  90.78 131.81    61.14   47.31  84.20 34.91
6 SCT_0001   1 2009  OH   0.98  92.99 141.71    64.88   51.72  90.97 37.33
   Avga  Avgb Avghue Avgchrom
1 29.28 27.58  43.19    40.40
2 27.14 25.18  43.44    37.32
3 27.46 25.55  43.39    37.81
4 26.26 23.53  42.95    35.55
5 26.60 24.64  43.91    36.66
6 28.75 25.53  42.46    38.73
>
```

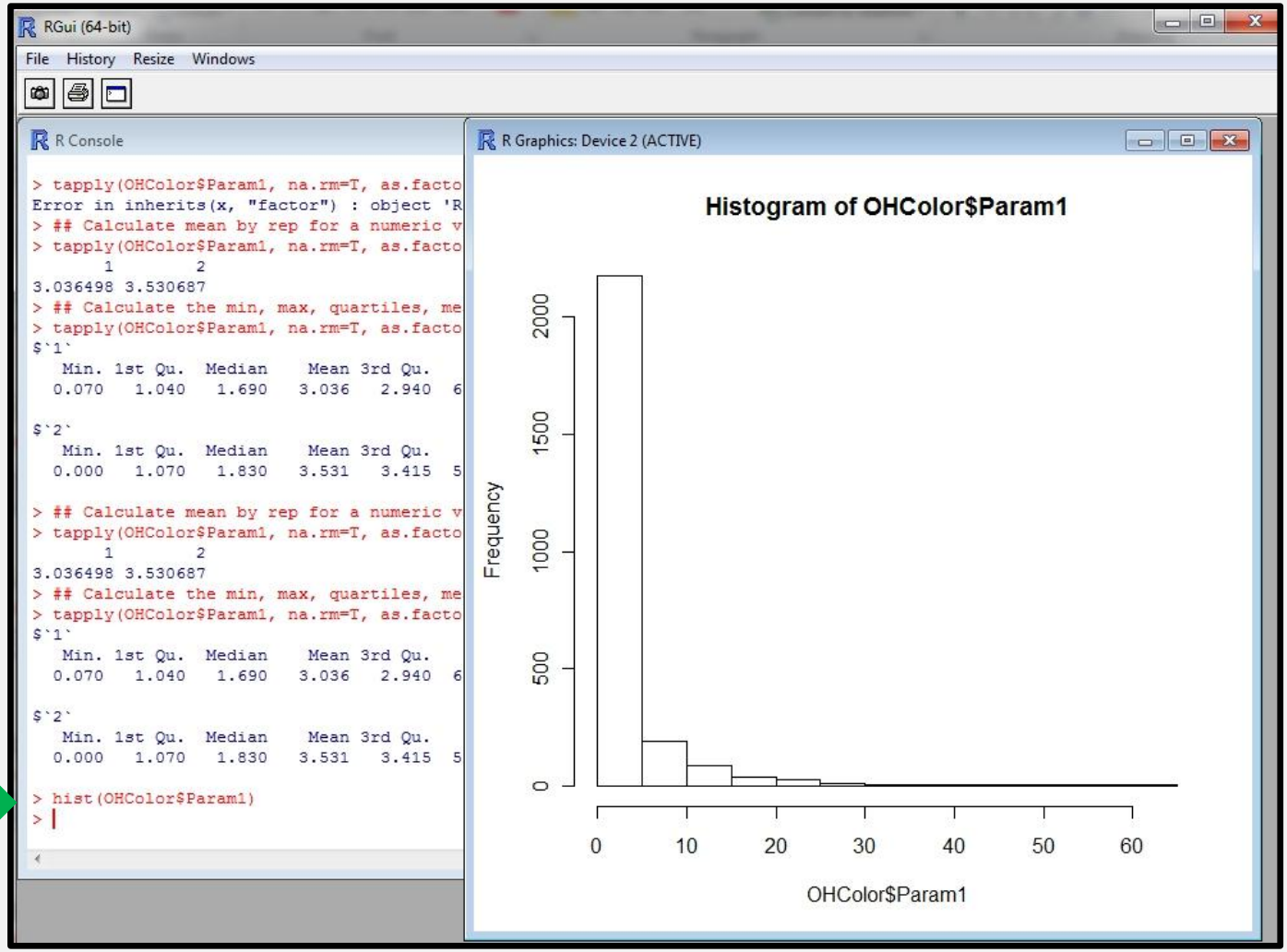
Checking Data – Structure Command

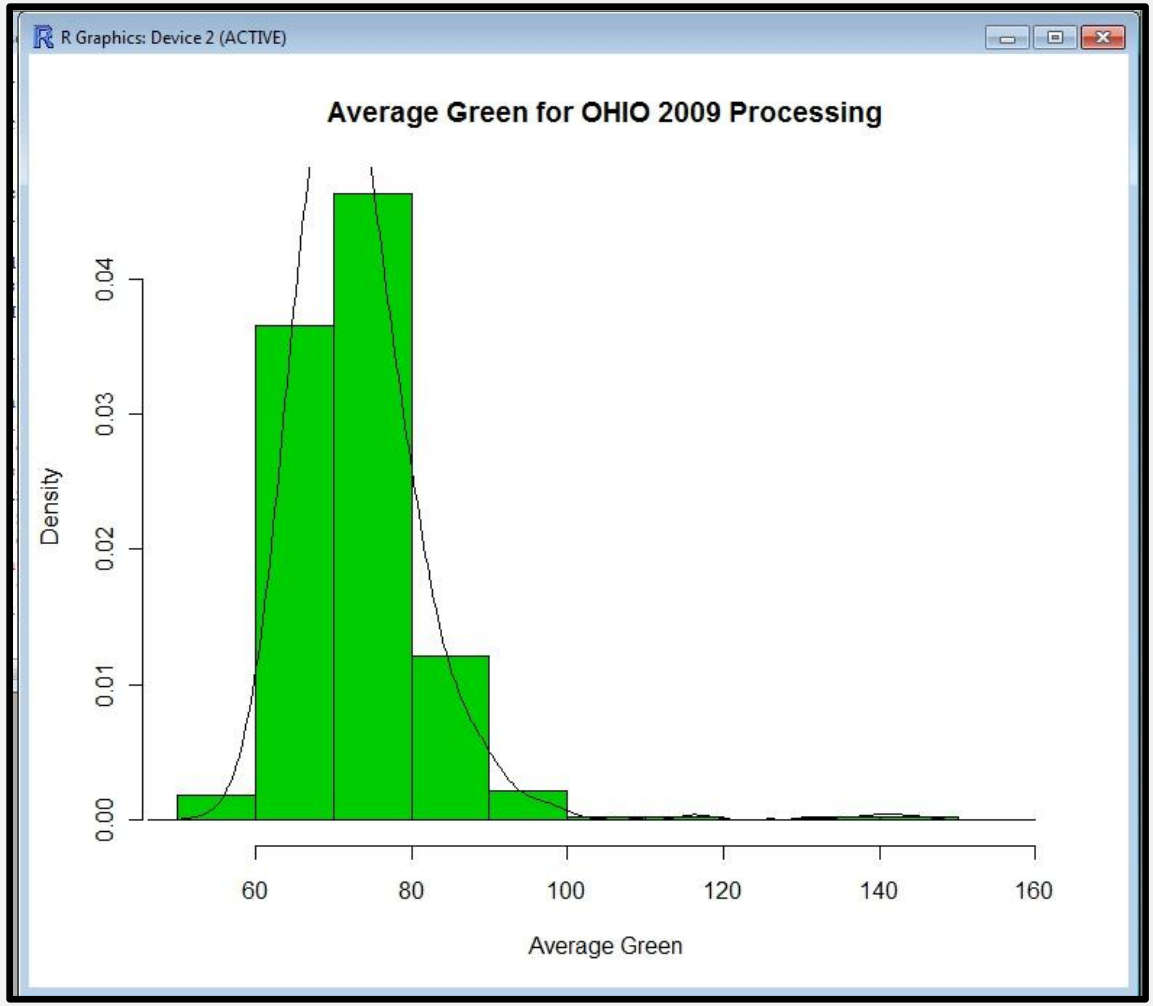
- str(object)
- See structure of the data set including the number of observations, number of variables, number of levels of categorical variables

```
> str(OHColor)
'data.frame':   2539 obs. of  15 variables:
 $ Line    : Factor w/ 143 levels "SCT_0001","SCT_0002",..: 1 1 1 1 1 1 1 1 1 2 ...
 $ Rep     : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Year    : int  2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...
 $ Loc     : Factor w/ 1 level "OH": 1 1 1 1 1 1 1 1 1 1 ...
 $ Param1  : num  0.81 1.55 1.17 1.49 1.26 0.98 0.47 0.62 1 0.99 ...
 $ Param2  : num  94.9 91.9 87.3 94.7 90.8 ...
 $ Avgred  : num  143 133 143 126 132 ...
 $ Avggreen: num  64.2 60.5 68.5 57 61.1 ...
 $ Avgblue : num  48.1 46.4 54 44.8 47.3 ...
 $ Avglum  : num  89.9 84.1 92.6 80.2 84.2 ...
 $ Avgl    : num  37.3 34.9 38.4 33 34.9 ...
 $ Avga    : num  29.3 27.1 27.5 26.3 26.6 ...
 $ Avgb    : num  27.6 25.2 25.6 23.5 24.6 ...
 $ Avghue  : num  43.2 43.4 43.4 43 43.9 ...
 $ Avgchrom: num  40.4 37.3 37.8 35.5 36.7 ...
>
```

Visualizing Data

- hist(object)






```
hist(OHColor$Avggreen, prob=T, xlab="Average Green",
ylab= "Density", main="Average Green for OHIO 2009
Processing", col=3); lines(density(OHColor$Avggreen,
na.rm=T, bw=2))
```

Learn More – Graphics

- Murrell, P. 2006. R graphics. Chapman & Hall/CRC, NY.

Simple ANOVA Models

- Are there differences in average green between lines?
- 2009OHColorSample.xls has color data from one year in one location
- $ access a subset of data
- lm(formula=model)
- anova(model)

```
> fit1 = lm(formula=OHColor$Avggreen~as.factor(OHColor$Line))
> anova(fit1)
Analysis of Variance Table

Response: OHColor$Avggreen
                          Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(OHColor$Line)  140 140844 1006.03  33.559 < 2.2e-16 ***
Residuals               2394  71768   29.98
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> |
```

Simplifying the R Code

- **Complicated code:**

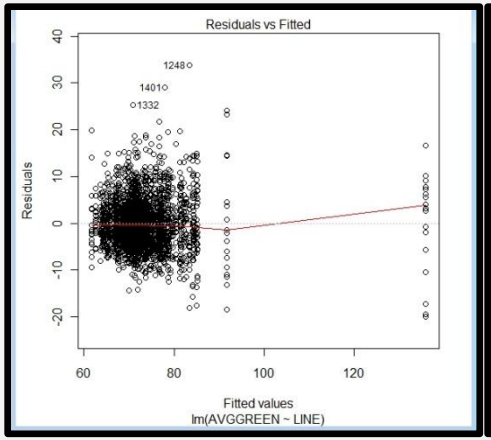
```
> fit1 = lm(formula=OHColor$Avggreen~as.factor(OHColor$Line))
> anova(fit1)
Analysis of Variance Table

Response: OHColor$Avggreen
                          Df Sum Sq Mean Sq F value    Pr(>F)
as.factor(OHColor$Line)  140 140844 1006.03  33.559 < 2.2e-16 ***
Residuals               2394  71768   29.98
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

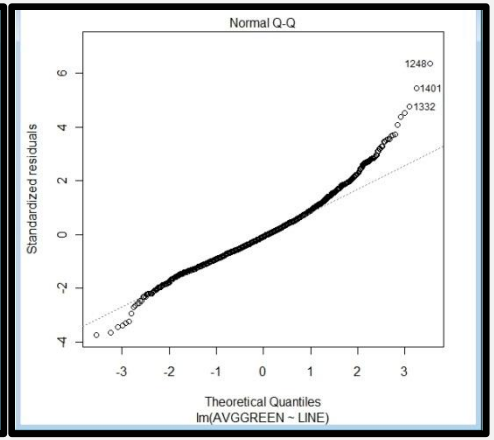
- **Simplified code:**

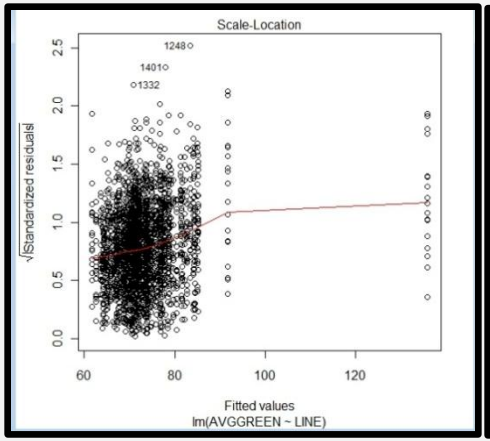
```
> # Rename variables for ease of use
> LINE = as.factor(OHColor$Line)
> AVGGREEN = OHColor$Avggreen
> fit1a = lm(AVGGREEN ~ LINE)
> anova(fit1a)
Analysis of Variance Table

Response: AVGGREEN
            Df Sum Sq Mean Sq F value    Pr(>F)
LINE       140 140844 1006.03  33.559 < 2.2e-16 ***
Residuals 2394  71768   29.98
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

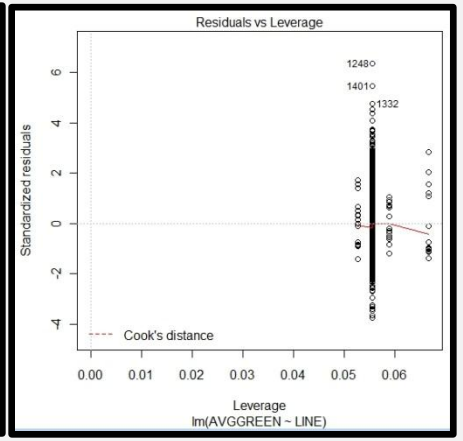
Test Assumptions

- plot(model)
- Provides 4 graphs: residuals vs. fits, qqplot, scale-location, residuals vs. leverage
- Meeting expected distributions challenging with large data sets









ANOVA Summary

- summary(model)
- Suggests multiple lines have mean average green level that differs from SCT_0001 (Intercept)
- Follow-up with t-tests, box plots (multiple comparisons not covered here)

```
> summary(fit1)

Call:
lm(formula = OHColor$Avggreen ~ as.factor(OHColor$Line))

Residuals:
    Min      1Q  Median      3Q     Max 
-19.923  -3.462  -0.422   2.829  33.811 

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                      62.5517     1.2905  48.470  < 2e-16 ***
as.factor(OHColor$Line)SCT_0002   6.1678     1.8251   3.379 0.000738 ***
as.factor(OHColor$Line)SCT_0003   8.8022     1.8251   4.823 1.50e-06 ***
as.factor(OHColor$Line)SCT_0004  18.8900     1.8251  10.350  < 2e-16 ***
as.factor(OHColor$Line)SCT_0005  10.9556     1.8251   6.003 2.23e-09 ***
as.factor(OHColor$Line)SCT_0006  29.2317     1.8251  16.017  < 2e-16 ***
as.factor(OHColor$Line)SCT_0007   8.3211     1.8251   4.559 5.39e-06 ***
as.factor(OHColor$Line)SCT_0008  12.0072     1.8251   6.579 5.80e-11 ***
as.factor(OHColor$Line)SCT_0009   7.4361     1.8251   4.074 4.76e-05 ***
as.factor(OHColor$Line)SCT_0010  12.2961     1.8251   6.737 2.01e-11 ***
as.factor(OHColor$Line)SCT_0011  14.7861     1.8251   8.102 8.55e-16 ***
as.factor(OHColor$Line)SCT_0012  22.6556     1.8251  12.413  < 2e-16 ***
as.factor(OHColor$Line)SCT_0013  14.2261     1.8251   7.795 9.54e-15 ***
as.factor(OHColor$Line)SCT_0014  18.7144     1.8251  10.254  < 2e-16 ***
as.factor(OHColor$Line)SCT_0015  15.4872     1.8251   8.486  < 2e-16 ***
as.factor(OHColor$Line)SCT_0016  22.5422     1.8251  12.351  < 2e-16 ***
as.factor(OHColor$Line)SCT_0017  20.1100     1.8251  11.019  < 2e-16 ***
```

Summarizing Data

- Prior to t-test, may want to summarize data
- mean(object)
- sd(object)
- na.rm=T remove missing data

```
> ## Calculate mean for a numeric variable and ignore missing data
> mean(AVGGREEN, na.rm=T)
[1] 73.17786
> ## Calculated standard deviation for a numeric variable and ignore missing data
> sd(AVGGREEN, na.rm=T)
[1] 9.15988
>
```

- **tapply - command used to apply a function, e.g. mean**

```
> ## Calculate mean by rep for a numeric variable
> tapply(AVGGREEN, na.rm=T, as.factor(OHColor$Rep), mean)
       1        2
72.65008 73.70606
> ## Calculate min, max, mean, median, first quartile, third quartile by rep for a numeric variable
> tapply(AVGGREEN, na.rm=T, as.factor(OHColor$Rep), summary)
$`1`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  52.27   67.65   71.68   72.65   76.07  145.00    2.00

$`2`
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
  55.55   67.38   72.42   73.71   77.87  152.50    2.00
```

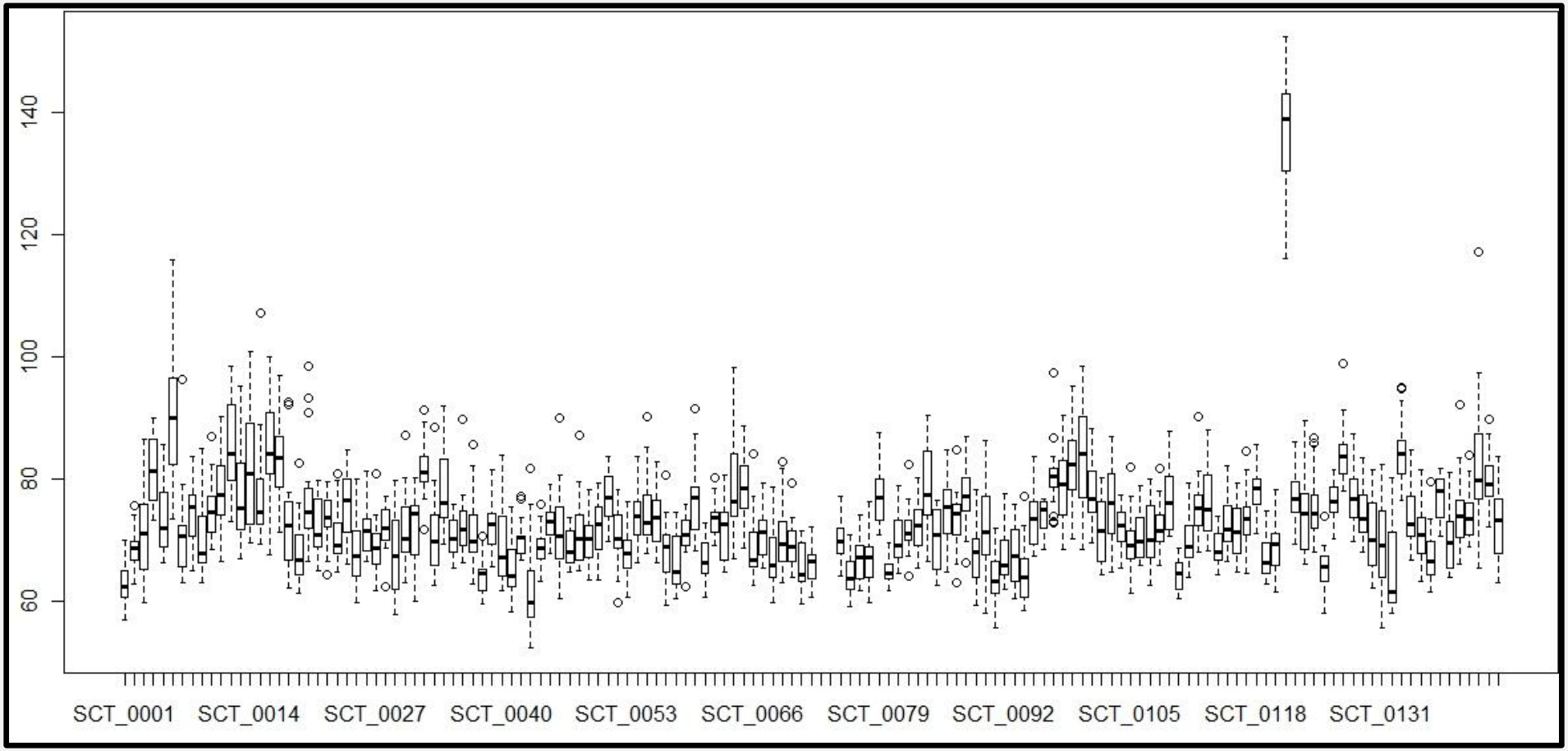
t-test

- Does a line have a higher average green value than the overall mean?
- Does one line have a higher average green than another?
- t.test(x,y) – x and y are numeric vectors
- Default confidence level is 0.95. Adjust by including (conf.level = *insert desired level*) in the command

```
> ## Does a line have a higher average green value than the overall mean?
> # t-test requires that we have numeric vectors
> # AVGGREEN is already a numeric vector
> # create a numeric vector with all average green values for a line
> sct0006<- OHColor[OHColor$Line == "SCT_0006", "Avggreen"]
> # check vector
> sct0006
 [1]  92.58  96.39  88.04  89.73  90.41  90.32  84.41  80.33  78.63  85.73  80.78  95.50 115.09 115.85
[15] 106.18  73.42 106.41  82.30
> # t-test
> t.test(sct0006, AVGGREEN, alternative="greater", var.equal=T)

        Two Sample t-test

data:  sct0006 and AVGGREEN
t = 8.5648, df = 2551, p-value < 2.2e-16
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 15.03104      Inf
sample estimates:
mean of x mean of y
 91.78333  73.17786
```

Box Plot

- boxplot(model)
- boxplot(AVGGREEN~LINE)



Combining Data Sets

- Combine data for 2009 and 2010
- 2010 data – 2010OHColorSample.xls
- Import 2010 data
- Dataframes need to have same headers
- rbind(dataframe1, dataframe2)

```
> ## Combine data from both years
> CombinedColor = rbind(OHColor, OHColor2010)
> head(CombinedColor)
      Line Rep Year Loc Param1 Param2 Avgred Avggreen Avgblue Avglum  Avgl
1 SCT_0001   1 2009  OH   0.81  94.87 142.86    64.24   48.14  89.87 37.34
2 SCT_0001   1 2009  OH   1.55  91.92 132.52    60.50   46.36  84.14 34.86
3 SCT_0001   1 2009  OH   1.17  87.32 142.92    68.52   54.04  92.62 38.35
4 SCT_0001   1 2009  OH   1.49  94.71 125.71    56.97   44.78  80.18 32.98
5 SCT_0001   1 2009  OH   1.26  90.78 131.81    61.14   47.31  84.20 34.91
6 SCT_0001   1 2009  OH   0.98  92.99 141.71    64.88   51.72  90.97 37.33
   Avga  Avgb Avghue Avgchrom
1 29.28 27.58  43.19    40.40
2 27.14 25.18  43.44    37.32
3 27.46 25.55  43.39    37.81
4 26.26 23.53  42.95    35.55
5 26.60 24.64  43.91    36.66
6 28.75 25.53  42.46    38.73
> tail(CombinedColor)
         Line Rep Year Loc Param1 Param2 Avgred Avggreen Avgblue Avglum  Avgl
3785 SCT_0478   2 2010  OH   1.04  34.25 140.28    73.43   46.19  87.74 37.91
```

Multi-year Data

- Sample data – collected in one location, two years, three reps total

```
> str(CombinedColor)
'data.frame':   3790 obs. of  15 variables:
 $ Line    : Factor w/ 143 levels "SCT_0001","SCT_0002",..: 1 1 1 1 1 1 1 1 1 2$
 $ Rep     : int  1 1 1 1 1 1 1 1 1 1 ...
 $ Year    : int  2009 2009 2009 2009 2009 2009 2009 2009 2009 2009 ...
 $ Loc     : Factor w/ 1 level "OH": 1 1 1 1 1 1 1 1 1 1 ...
 $ Param1  : num  0.81 1.55 1.17 1.49 1.26 0.98 0.47 0.62 1 0.99 ...
 $ Param2  : num  94.9 91.9 87.3 94.7 90.8 ...
 $ Avgred  : num  143 133 143 126 132 ...
 $ Avggreen: num  64.2 60.5 68.5 57 61.1 ...
 $ Avgblue : num  48.1 46.4 54 44.8 47.3 ...
 $ Avglum  : num  89.9 84.1 92.6 80.2 84.2 ...
 $ Avgl    : num  37.3 34.9 38.4 33 34.9 ...
 $ Avga    : num  29.3 27.1 27.5 26.3 26.6 ...
 $ Avgb    : num  27.6 25.2 25.6 23.5 24.6 ...
 $ Avghue  : num  43.2 43.4 43.4 43 43.9 ...
 $ Avgchrom: num  40.4 37.3 37.8 35.5 36.7 ...
>
```

Assign Variable Names

- Rename variables so that rep and year are factors for ease of use
- R recognizes the most recent object name if the name is used multiple times (e.g. we previously assigned the name AVGGREEN to average green in the 2009 data. AVGGREEN is now assigned to the combined data)

```
> LINE=as.factor(CombinedOHColor$Line)
> REP=as.factor(CombinedOHColor$Rep)
> YEAR=as.factor(CombinedOHColor$Year)
> AVGGREEN=as.numeric(CombinedOHColor$Avggreen)
>
```

Multi-year ANOVA

- REP and YEAR considered fixed
- Denote nesting using %in%
- Denote interactions using a colon or asterix between terms

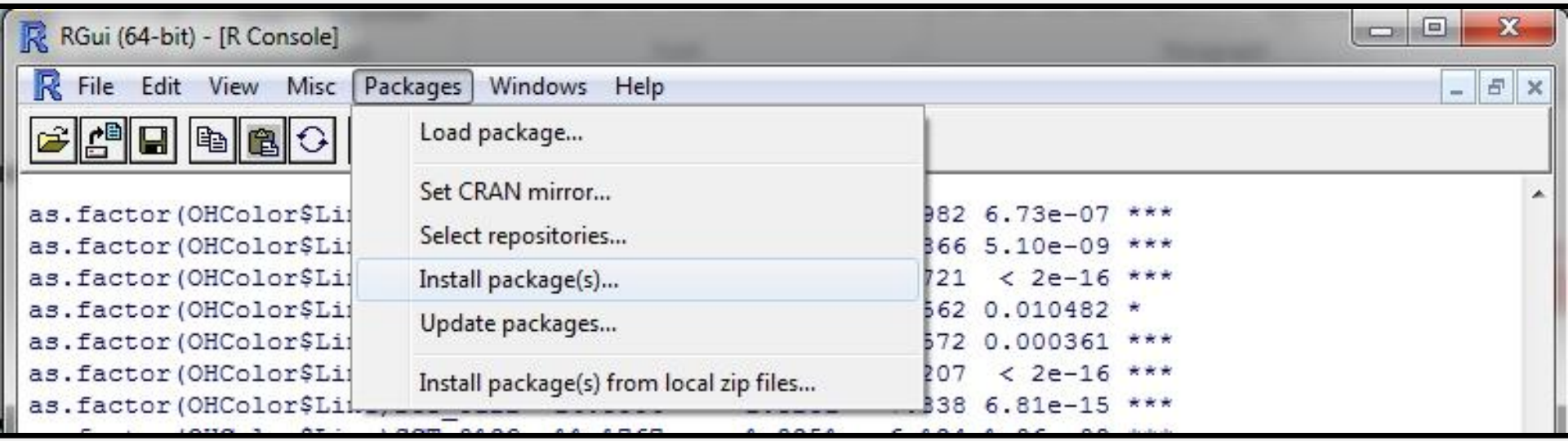
```
> # Create and test model
> fit2 =lm(AVGGREEN~ LINE + YEAR + REP%in%YEAR + LINE:YEAR)
> anova(fit2)
Analysis of Variance Table

Response: AVGGREEN
            Df Sum Sq Mean Sq F value    Pr(>F)
LINE       140 203748 1455.34 49.4371 < 2.2e-16 ***
YEAR         1    429  429.24 14.5811 0.0001366 ***
YEAR:REP     1    707  706.56 24.0013 1.006e-06 ***
LINE:YEAR  137   9606   70.12  2.3819 < 2.2e-16 ***
Residuals 3506 103211   29.44
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
```

Models with Random Effects

- For many plant breeding applications we would consider main effects to be random, and would want to estimate the proportion of variance due to effects in our experimental design (e.g. estimates of heritability)
- Requires lme4 package

Installing a Package

- First time you want to use a package



loading a Package

- Load package every R session you want to use it
- library(package name)

```
> library(lme4)
Loading required package: Matrix
Loading required package: lattice

Attaching package: 'Matrix'

The following object(s) are masked from 'package:base':

    det


Attaching package: 'lme4'

The following object(s) are masked from 'package:stats':

    AIC

Warning messages:
1: package 'lme4' was built under R version 2.12.2
2: package 'Matrix' was built under R version 2.12.2
>
```

Calculating Variance Components

- lmer(model)
- Denote random effect - (1|object)
- lmer can also be used with a mixed model

```
> # Create model
> fit3 = lmer(AVGGREEN~(1|LINE) + (1|YEAR) + (1|REP%in%YEAR) + (1|LINE:YEAR))
> summary(fit3)
Linear mixed model fit by REML
Formula: AVGGREEN ~ (1 | LINE) + (1 | YEAR) + (1 | REP %in% YEAR) + (1 |      LINE:YEAR)
   AIC   BIC logLik deviance REMLdev
 24253 24291 -12121    24243   24241
Random effects:
 Groups        Name        Variance  Std.Dev.
 LINE:YEAR     (Intercept)  3.340655 1.82775
 LINE          (Intercept) 51.395213 7.16905
 YEAR          (Intercept)  0.204915 0.45267
 REP %in% YEAR (Intercept)  0.021804 0.14766
 Residual                  29.630945 5.44343
Number of obs: 3786, groups: LINE:YEAR, 279; LINE, 141; YEAR, 2; REP %in% YEAR, 1

Fixed effects:
            Estimate Std. Error t value
(Intercept)  73.5083     0.7128   103.1
>
```

Can be used to estimate heritability

Learn More – lme4 Package

- Bates, D., M. Maechler, and B. Bolker. 2011. Package 'lme4' [Online]. The R Project for Statistical Computing. Available at: http://cran.r-project.org/web/packages/lme4/lme4.pdf (verified 13 Sept 2011).

Exporting Tables

- Useful after manipulating a data set or creating a new data set
- Combined data for two years
- write.table(dataframe, col.names=NA, “filename.txt”)

```
> ## Exporting a table
> # For PC
> write.table(CombinedColor, col.names=NA, "C:/Users/merk.9/Desktop/Documents/2011 Webinar Series/CombinedColorData.txt")
> # For Mac
> write.table(CombinedColor, col.names=NA, "/Users/heathermerk/Documents/eXtension/2011 Webinar Series/CombinedColorData.txt")
```

Introduction to loops

- Based on – Cock, P. Programming in R [Online]. Molecular Organization and Assembly in Cells, The University of Warwick. Available at: http://www2.warwick.ac.uk/fac/sci/moac/degrees/modules/ch923/r_introduction/r_programming (verified 14 Sept 2011).
- Save time and simplify code
- for and while loops

For loops

- Take the form
  - for (variable in sequence) expression

  OR
  - for (variable in sequence)

  {

  expression

  expression

  expression

  }

```
> for(x in c(1:10)) print(sqrt(x))
[1] 1
[1] 1.414214
[1] 1.732051
[1] 2
[1] 2.236068
[1] 2.449490
[1] 2.645751
[1] 2.828427
[1] 3
[1] 3.162278
> for (x in c(1:10))
+ {
+ print(sqrt(x))
+ }
[1] 1
[1] 1.414214
[1] 1.732051
[1] 2
[1] 2.236068
[1] 2.449490
[1] 2.645751
[1] 2.828427
[1] 3
[1] 3.162278
```

While loops

- Take the form
  - while(condition) expression

  OR

  - while(condition)

  {

  expression

  expression

  expression

  }

```
> ## while loop - Fibonacci series
> ## each number is the sum of the
> ## previous two numbers
> a<-0
> b<-1
> print(a)
[1] 0
> while(b<50)
+ {
+ print(b)
+ temp<-a+b
+ a<-b
+ b<-temp
+ }
[1] 1
[1] 1
[1] 2
[1] 3
[1] 5
[1] 8
[1] 13
[1] 21
[1] 34
```

Single Marker-Trait Analysis

- Test association between trait and marker, one marker at a time
- Use simple linear model, lm(trait~marker)

	A	B	C	D	E
1	Trait	Line	Marker 1	Marker 2	Marker 3
2	35.855	2k1-1439	A	T	A
3	37.608	2k1-2019	T	T	A
4	38.732	21k-2020	T	T	A
5	41.996	21k-2054	T	T	A
6	39.603	CULBPT04	A	T	A

Single Marker-Trait Analysis loop

```
for(x in 3:ncol(data))
{
print(names(data[x]))
print(anova(lm(trait~data[,x])))
}
```

R Help

- ??function name
- help(function name)
- Help menu option of GUI

R Help Mailing List

https://stat.ethz.ch/mailman/listinfo/r-help

R-help -- Main R Mailing List: Primary help

About R-help

View this page in
English (USA)

The main R mailing list, for announcements about the development of R and the availability of new code, questions and answers about problems and solutions using R, enhancements and patches to the source code and documentation of R, comparison and compatibility with S and S-plus, and for the posting of nice examples and benchmarks. Please read the *General Instructions* on the R Mailing Lists page **and** follow the ***posting guide***!

Learn More – Online

www.eXtension.org/plant_breeding_genomics

R content coming Fall 2011

Home | About | Resource Areas | News | Articles | Answers | Calendar | Learning Lessons | Search

Plant Breeding and Genomics

Here are some of our featured articles and activities...

Conifer Translational Genomics Network Online Modules

A series of 16 forest tree breeding online learning modules

More...

1 2 3 4 5 6 7 8 Next

In This Resource Area

Plant Breeding and Genomics Topics

- Analysis with SAS
- Data Sets
- Experimental Design and Statistical Theory

Answers from our Experts

September 08, 2011

I'd like to screen potato varieties for disease resistance markers. Is there a list of known markers...

This resource area was created by the: Plant Breeding and Genomics community

In The News...

September 06, 2011
Computer Helps Michigan State University Researchers Unravel Plants' Secrets to Survival

September 05, 2011
Free September Webinars Focus on Entrepreneurs, Feed for Dairy Cattle and Horses, Plant Breeding, Manure, and Facilities for Meat Processing

August 22, 2011
SolRgene: an Online Database to Explore Disease Resistance Genes in Tuber-bearing Solanum Species

More ...

Resource Area Feeds

- Track all new content

Learn More – Online

- Kim, D.Y. R basics [Online]. Illinois State University. Available at: http://math.illinoisstate.edu/dhkim/rstuff/rtutor.html (verified 8 Sept 2011).

- Martinez, M. R for biologists [Online]. The R Project for Statistical Computing. Available at: http://cran.r-project.org/doc/contrib/Martinez-RforBiologistv1.1.pdf (verified 8 Sept 2011).

- Verzani, J. SimpleR – Using R for introductory statistics [Online]. The R Project for Statistical Computing. Available at: http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf (verified 8 Sept 2011).

Learn More – In Person

- Summer Institute for Statistical Genetics (SISG) at the University of Washington - http://www.biostat.washington.edu/suminst/sisg/general

- useR! Conference - http://www.r-project.org/conferences.html

Learn More – General Texts

- Adler, J. 2010. R in a nutshell: a desktop quick reference. O'Reilly, Sebastapol, CA.
- Crawley, M. 2007. The R book. Wiley, Hoboken, NJ.
- Dalgaard, P. 2008. Introductory statistics with R. Springer-Verlaugh, NY.
- Zuur, A. F., E. N. Leno, and E.H.W.G. Meesters. 2009. A beginner's guide to R. Springer, NY.

Learn More – Importing Data

- R Development Core Team. R Data import/export [Online]. The Comprehensive R Archive Network. Available at: http://cran.r-project.org (verified 9 Sept 2011).

Learn More – Graphics

- Murrell, P. 2006. R graphics. Chapman & Hall/CRC, NY.

Learn More – lme4 Package

- Bates, D., M. Maechler, and B. Bolker. 2011. Package 'lme4' [Online]. The R Project for Statistical Computing. Available at: http://cran.r-project.org/web/packages/lme4/lme4.pdf (verified 13 Sept 2011).

Learn More – Programming in R

- Cock, P. Programming in R [Online]. Molecular Organization and Assembly in Cells, The University of Warwick. Available at: http://www2.warwick.ac.uk/fac/sci/moac/degrees/modules/ch923/r_introduction/r_programming (verified 14 Sept 2011).

Acknowledgements

- David Francis, The Ohio State University
- Debora Liabeuf, The Ohio State University
- Sung-Chur Sim, The Ohio State University
- Walter De Jong, Cornell University

- John McQueen, Oregon State University – Technical Support
- Michael Coe, Cedar Lake Research Group - Evaluator

Supplemental Files

- Color data collected in Ohio in 2009
  - 2009OHColorSample.xls
- Color data collected in Ohio in 2010
  - 2010OHColorSample.xls
- Script file
  - Script2009OHColorSample.txt
- All files available at: http://www.extension.org/pages/60427/

Please fill out the survey evaluation! (You will be contacted via email)