U3\_software.R

hirsc

Tue Oct 16 20:15:59 2018

library(tidyverse)

## -- Attaching packages ------------------------------ tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts --------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(gplots)

##   
## Attaching package: 'gplots'

## The following object is masked from 'package:stats':  
##   
## lowess

library(mosaic)

## Loading required package: lattice

## Loading required package: ggformula

## Loading required package: ggstance

##   
## Attaching package: 'ggstance'

## The following objects are masked from 'package:ggplot2':  
##   
## geom\_errorbarh, GeomErrorbarh

##   
## New to ggformula? Try the tutorials:   
## learnr::run\_tutorial("introduction", package = "ggformula")  
## learnr::run\_tutorial("refining", package = "ggformula")

## Loading required package: mosaicData

## Loading required package: Matrix

##   
## Attaching package: 'Matrix'

## The following object is masked from 'package:tidyr':  
##   
## expand

##   
## The 'mosaic' package masks several functions from core packages in order to add   
## additional features. The original behavior of these functions should not be affected by this.  
##   
## Note: If you use the Matrix package, be sure to load it BEFORE loading mosaic.

##   
## Attaching package: 'mosaic'

## The following object is masked from 'package:Matrix':  
##   
## mean

## The following objects are masked from 'package:dplyr':  
##   
## count, do, tally

## The following object is masked from 'package:purrr':  
##   
## cross

## The following object is masked from 'package:ggplot2':  
##   
## stat

## The following objects are masked from 'package:stats':  
##   
## binom.test, cor, cor.test, cov, fivenum, IQR, median,  
## prop.test, quantile, sd, t.test, var

## The following objects are masked from 'package:base':  
##   
## max, mean, min, prod, range, sample, sum

library(janitor)  
library(agricolae)  
library(RcmdrMisc)

## Loading required package: car

## Loading required package: carData

##   
## Attaching package: 'car'

## The following objects are masked from 'package:mosaic':  
##   
## deltaMethod, logit

## The following object is masked from 'package:dplyr':  
##   
## recode

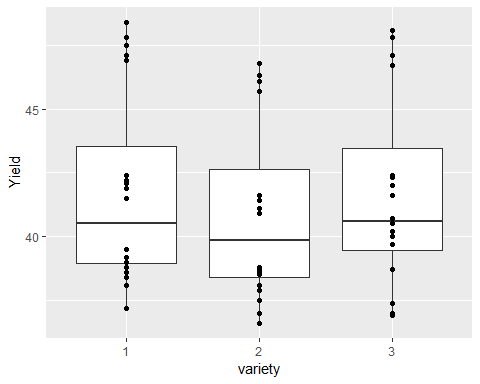
## The following object is masked from 'package:purrr':  
##   
## some

## Loading required package: sandwich

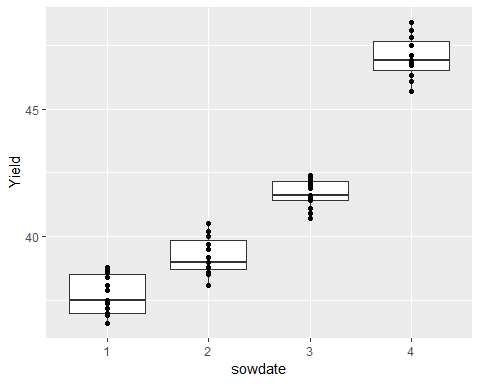
#2A  
soy <- read\_csv("snapbean.csv")

## Parsed with column specification:  
## cols(  
## sowdate = col\_integer(),  
## variety = col\_integer(),  
## rep = col\_integer(),  
## Earliness = col\_double(),  
## LeafAreaEarliness = col\_double(),  
## Yield = col\_double(),  
## AvgLeafArea = col\_integer()  
## )

soy$sowdate <- as.factor(soy$sowdate)  
soy$variety <- as.factor(soy$variety)  
  
soy %>%   
 ggplot(aes(x = variety, y = Yield))+  
 geom\_boxplot()+  
 geom\_point()



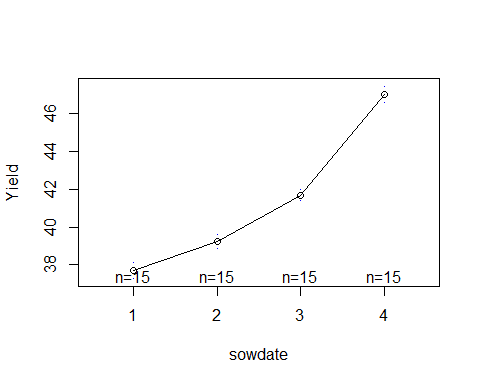
soy %>%   
 ggplot(aes(x = sowdate, y = Yield))+  
 geom\_boxplot()+  
 geom\_point()



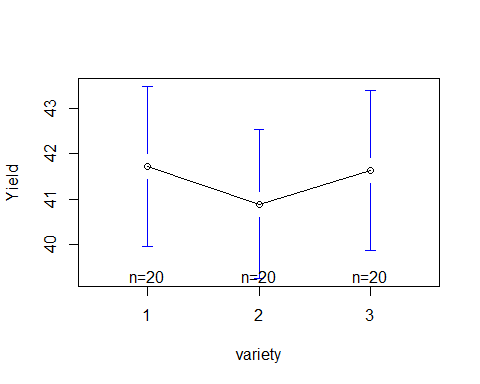
plotmeans(Yield~sowdate, data = soy)

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped

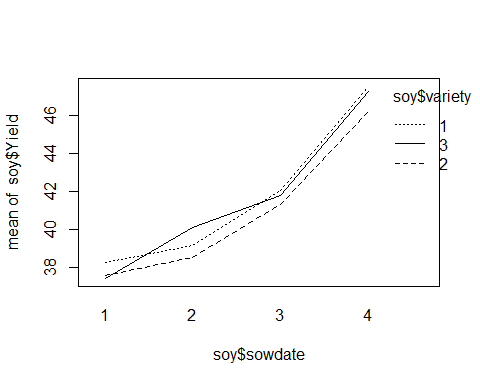
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, ui, x, pmin(y + gap, ui), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped



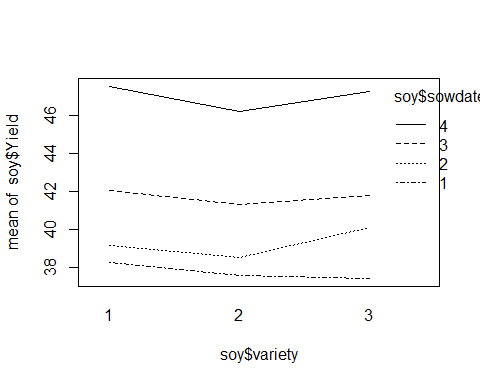
plotmeans(Yield~variety, data = soy)



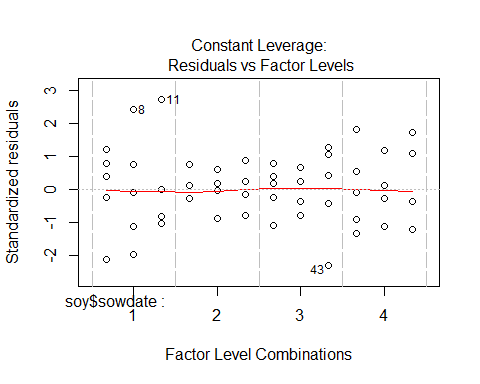
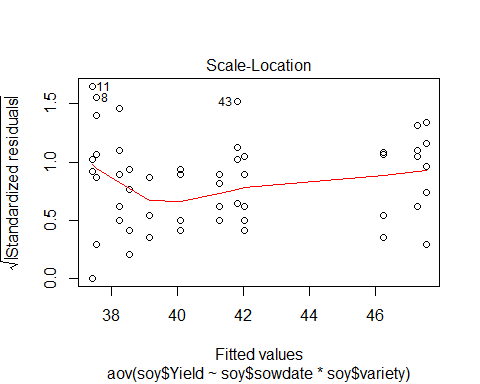
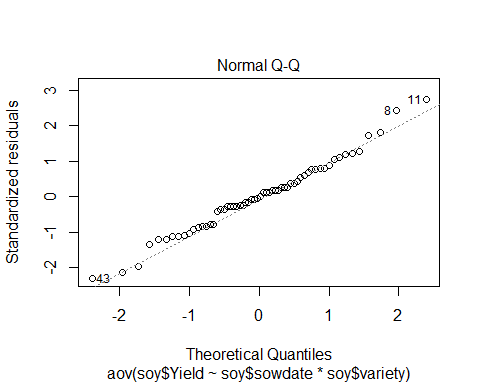
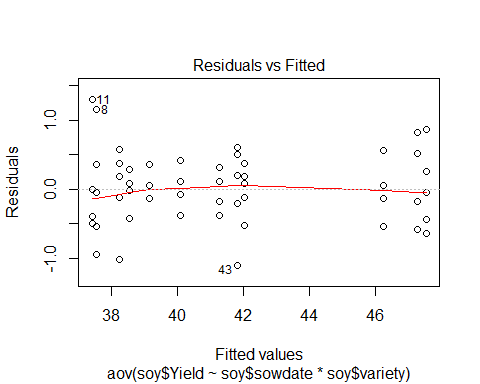
interaction.plot(soy$sowdate, soy$variety, soy$Yield)



interaction.plot(soy$variety, soy$sowdate, soy$Yield)



#2B  
  
my\_aov <- aov(soy$Yield~soy$sowdate\*soy$variety)  
plot(my\_aov)



numSummary(soy$Yield, groups=soy$sowdate)

## mean sd IQR 0% 25% 50% 75% 100% data:n  
## 1 37.72000 0.7747811 1.50 36.6 37.0 37.5 38.50 38.8 15  
## 2 39.24667 0.7059610 1.15 38.1 38.7 39.0 39.85 40.5 15  
## 3 41.70000 0.5424811 0.75 40.7 41.4 41.6 42.15 42.4 15  
## 4 47.02000 0.7757024 1.15 45.7 46.5 46.9 47.65 48.4 15

numSummary(soy$Yield, groups=soy$variety)

## mean sd IQR 0% 25% 50% 75% 100% data:n  
## 1 41.730 3.755571 4.575 37.2 38.95 40.50 43.525 48.4 20  
## 2 40.895 3.493825 4.225 36.6 38.40 39.85 42.625 46.8 20  
## 3 41.640 3.751407 4.025 36.9 39.45 40.60 43.475 48.1 20

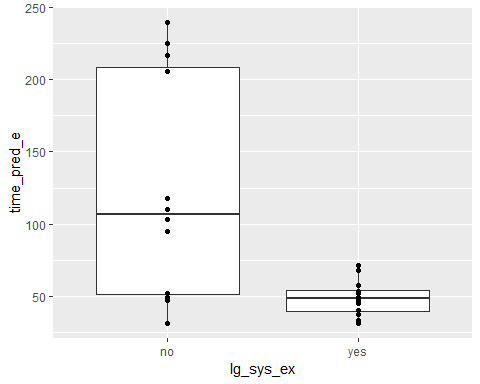
#2C  
anova(my\_aov)

## Analysis of Variance Table  
##   
## Response: soy$Yield  
## Df Sum Sq Mean Sq F value Pr(>F)   
## soy$sowdate 3 747.78 249.259 876.130 < 2.2e-16 \*\*\*  
## soy$variety 2 8.40 4.201 14.767 1.005e-05 \*\*\*  
## soy$sowdate:soy$variety 6 5.87 0.978 3.437 0.006645 \*\*   
## Residuals 48 13.66 0.284   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

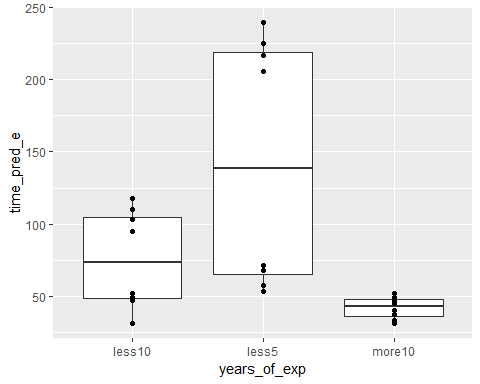
#3  
dat <- read\_csv("programmers.csv")

## Parsed with column specification:  
## cols(  
## LgSysEx = col\_character(),  
## YearsOfExp = col\_character(),  
## TimePredE = col\_integer()  
## )

dat <- clean\_names(dat)  
  
#3A  
dat%>%   
 ggplot(aes(x = lg\_sys\_ex, y = time\_pred\_e))+  
 geom\_boxplot()+  
 geom\_point()

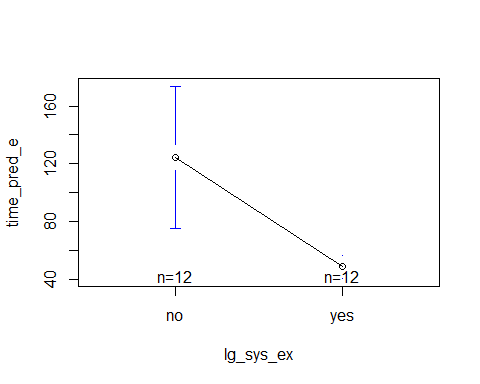


dat%>%   
 ggplot(aes(x = years\_of\_exp, y = time\_pred\_e))+  
 geom\_boxplot()+  
 geom\_point()



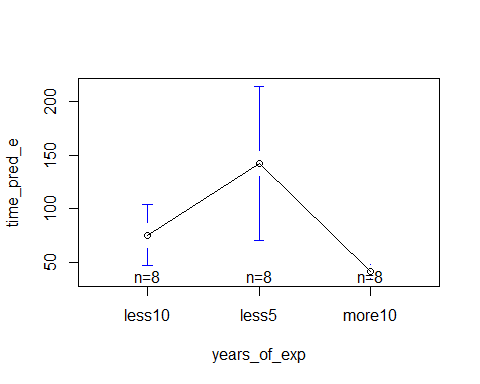
plotmeans(time\_pred\_e~lg\_sys\_ex, data = dat)

## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped

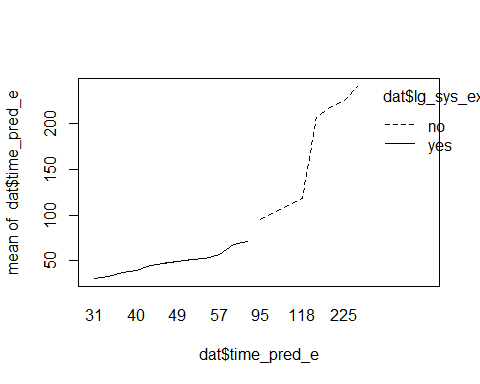


plotmeans(time\_pred\_e~years\_of\_exp, data = dat)

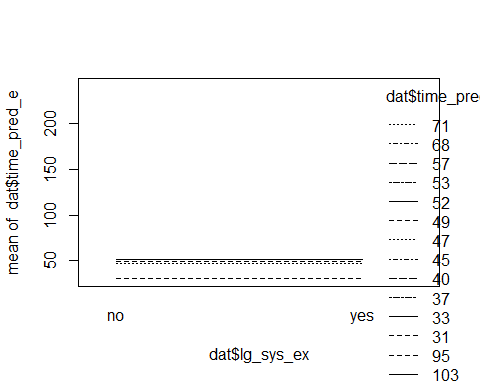
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped  
  
## Warning in arrows(x, li, x, pmax(y - gap, li), col = barcol, lwd = lwd, :  
## zero-length arrow is of indeterminate angle and so skipped



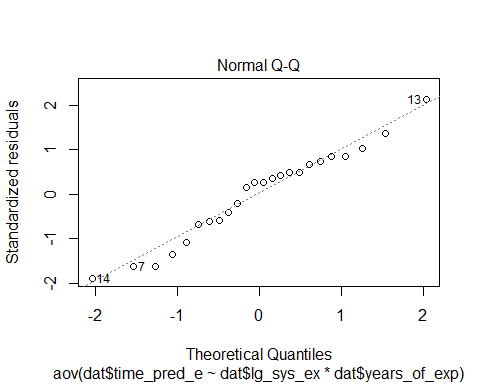
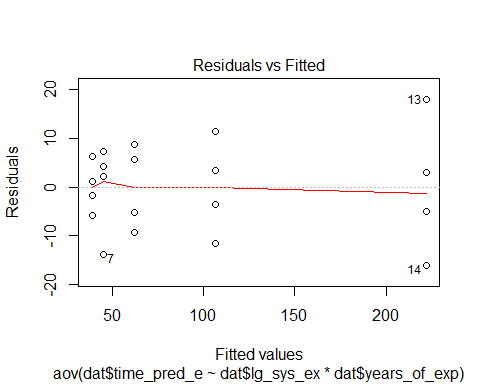
interaction.plot(dat$time\_pred\_e, dat$lg\_sys\_ex, dat$time\_pred\_e)



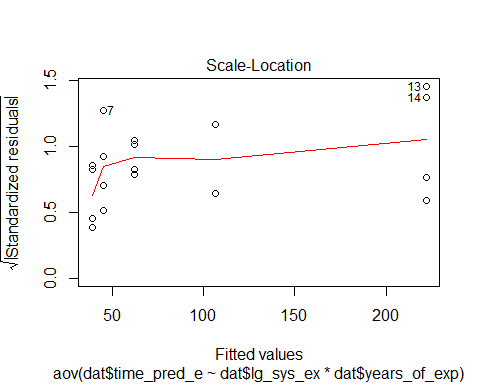
interaction.plot(dat$lg\_sys\_ex, dat$time\_pred\_e, dat$time\_pred\_e)



#3B  
my\_aov1 <- aov(dat$time\_pred\_e~dat$lg\_sys\_ex\*dat$years\_of\_exp)  
plot(my\_aov1)



## hat values (leverages) are all = 0.25  
## and there are no factor predictors; no plot no. 5



numSummary(dat$time\_pred\_e, groups=dat$lg\_sys\_ex)

## mean sd IQR 0% 25% 50% 75% 100% data:n  
## no 124.41667 77.41794 157.50 31 51.25 106.5 208.75 240 12  
## yes 48.58333 12.63803 14.75 31 39.25 48.0 54.00 71 12

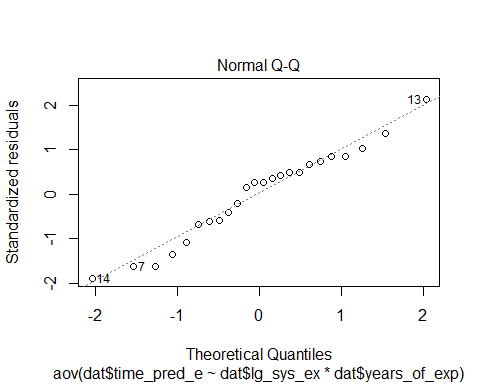
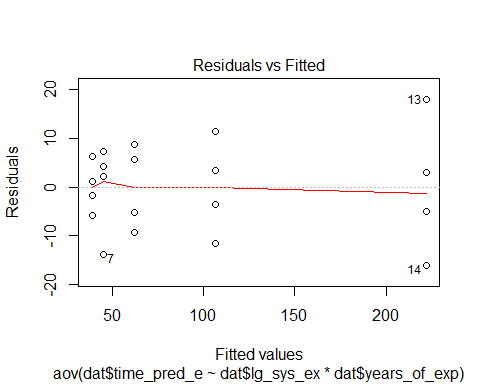
numSummary(dat$time\_pred\_e, groups=dat$years\_of\_exp)

## mean sd IQR 0% 25% 50% 75% 100% data:n  
## less10 75.625 34.184113 56.25 31 48.50 73.5 104.75 118 8  
## less5 142.125 86.087061 153.75 53 65.25 138.5 219.00 240 8  
## more10 41.750 7.685794 11.50 31 36.00 42.5 47.50 52 8

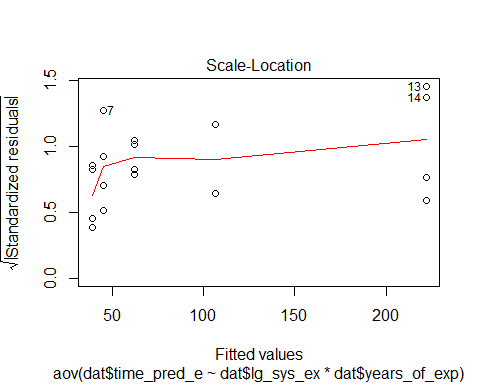
#3C  
anova(my\_aov1)

## Analysis of Variance Table  
##   
## Response: dat$time\_pred\_e  
## Df Sum Sq Mean Sq F value Pr(>F)   
## dat$lg\_sys\_ex 1 34504 34504 358.59 2.469e-13 \*\*\*  
## dat$years\_of\_exp 2 41720 20860 216.79 2.540e-13 \*\*\*  
## dat$lg\_sys\_ex:dat$years\_of\_exp 2 24234 12117 125.93 2.614e-11 \*\*\*  
## Residuals 18 1732 96   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

#4  
y\_aov1 <- aov(log(dat$time\_pred\_e)~dat$lg\_sys\_ex\*dat$years\_of\_exp)  
plot(my\_aov1)



## hat values (leverages) are all = 0.25  
## and there are no factor predictors; no plot no. 5



#numSummary(dat$time\_pred\_e["Logtime"], groups=dat$lg\_sys\_ex)  
#numSummary(dat$time\_pred\_e("log"), groups=dat$years\_of\_exp)  
  
anova(y\_aov1)

## Analysis of Variance Table  
##   
## Response: log(dat$time\_pred\_e)  
## Df Sum Sq Mean Sq F value Pr(>F)   
## dat$lg\_sys\_ex 1 3.5005 3.5005 131.091 1.071e-09 \*\*\*  
## dat$years\_of\_exp 2 4.3798 2.1899 82.010 9.044e-10 \*\*\*  
## dat$lg\_sys\_ex:dat$years\_of\_exp 2 1.3565 0.6782 25.399 5.744e-06 \*\*\*  
## Residuals 18 0.4806 0.0267   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1