ΤΙΜΕΣ ΔΗΜΗΤΡΙΑΚΩΝ ΚΑΙ Η ΑΛΛΗΛΕΠΙΔΡΑΣΗ ΤΟΥΣ

<https://www.kaggle.com/datasets/timmofeyy/-cerial-prices-changes-within-last-30-years?select=rice_wheat_corn_prices.csv>

Από τον παραπάνω σύνδεσμο παίρνουμε την βάση δεδομένων με τις τιμές των δημητριακών για τα τελευταία 30 χρονιά.

table <- read.csv("C:/Users/john/Desktop/rice\_wheat\_corn\_prices.csv")

Οι τιμές αντιστοιχούν σε έναν τόνο του εκάστοτε δημητριακού

Ομαδοποιούμε τα δεδομένα μας σε ετήσια βάση

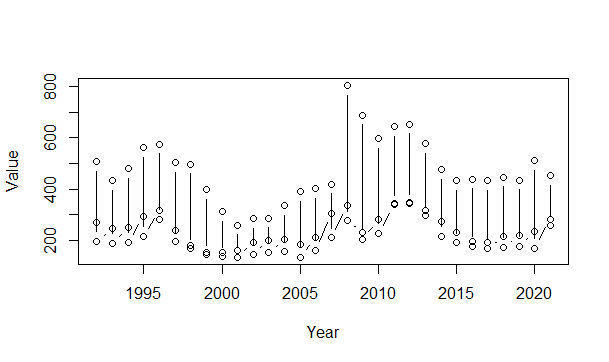
yearly\_price <- table %>% filter(Year != 2022) %>% group\_by(Year) %>% summarise\_at(vars('Price\_wheat\_ton\_infl','Price\_rice\_ton\_infl','Price\_corn\_ton\_infl'),mean)

και συγκεντρώνουμε τα αποτελέσματα μας σε 3 πεδία για καλύτερη οπτικοποιηση αποτελεσμάτων

cereal\_category <- yearly\_price %>% pivot\_longer(2:4, names\_to = 'cereal')

δημιουργούμε το διάγραμμα

plot(cereal\_category$Year,cereal\_category$value, type = "b", xlab = "Year", ylab = "Value")



Εδώ παρατηρούμε ότι οι αυξομειώσεις και των 3 ειδών δημητριακών είναι ίδιες

Επίσης παρατηρούμε ότι οι τιμές των δημητριακών έχουν μεγάλες διακυμάνσεις δεν υπάρχει σταθερότητα.

Μετα θέλουμε να δούμε τις τιμές των δημητριακών για τις χρονιές όπου το σιτάρι κόστιζε πάνω από 300$

high\_wheat\_price <- yearly\_price[yearly\_price$Price\_wheat\_ton\_infl >= 300, c(1,2,3,4)]

summary(high\_wheat\_price)

Year Price\_wheat\_ton\_infl Price\_rice\_ton\_infl Price\_corn\_ton\_infl

Min. :1996 Min. :306.1 Min. :418.8 Min. :210.0

1st Qu.:2007 1st Qu.:316.5 1st Qu.:575.5 1st Qu.:277.1

Median :2010 Median :326.7 Median :610.0 Median :288.7

Mean :2008 Mean :326.0 Mean :611.5 Mean :292.3

3rd Qu.:2012 3rd Qu.:337.5 3rd Qu.:649.8 3rd Qu.:332.8

Max. :2013 Max. :342.2 Max. :803.4 Max. :345.8

**Με την βοήθεια των μοντέλων παλινδρόμησης μπορούμε να προβλέψουμε τιμές για κατηγορίες σιτηρών εάν γνωρίζουμε την τιμή μιας άλλης κατηγορίας , εφόσον έχουμε υψηλό συντελεστή γραμμικής συσχέτισης.**

Για να προβλέψουμε την τιμή του σιταριού

Δημιουργούμε δυο μοντέλα παλινδρόμησης :

σιτάρι ~ ρύζι

model\_W\_R <- lm(yearly\_price$Price\_wheat\_ton\_infl~yearly\_price$Price\_rice\_ton\_infl)

Multiple R-squared: 0.5705

Ο συντελεστής γραμμικής συσχέτισης είναι αρκετά χαμηλός

coef(model\_W\_R)

(Intercept) yearly\_price$Price\_rice\_ton\_infl

80.0335550 0.3412051

και σιτάρι ~ καλαμπόκι

model\_W\_C <- lm(yearly\_price$Price\_wheat\_ton\_infl~yearly\_price$Price\_corn\_ton\_infl)

Multiple R-squared: 0.8396

Υψηλότερος βαθμός συσχέτισης μεταξύ τιμών σιταριού και καλαμποκιού οπότε το μοντέλο σιτάρι ~ καλαμπόκι είναι καλύτερο

coef(model\_W\_C)

(Intercept) yearly\_price$Price\_corn\_ton\_infl

62.0378370 0.8935009

Υπ1 : Η τιμή του ρυζιού φτάσει τα 800$ τότε σύμφωνα με το μοντέλο

W\_R\_predict = 80.0335 + 0.3412\*800

η τιμή του σιταριού θα καθοριστεί 352,9935$

Υπ2 : Η τιμή του καλαμποκιού φτάσει τα 400$ τότε σύμφωνα με το μοντέλο

W\_c\_predict = 62.0378 + 0.8935\*400

η τιμή του σιταριού θα καθοριστεί 419,4378$

**Ποιες ήταν οι τιμές για ρύζι και καλαμπόκι την χρονιά που το σιτάρι είχε την χαμηλότερη τιμή για τα τελευταία 30 χρονιά**

yearly\_price %>% select(Year, Price\_wheat\_ton\_infl, Price\_rice\_ton\_infl,Price\_corn\_ton\_infl) %>% slice\_min(yearly\_price$Price\_wheat\_ton\_infl)

Year Price\_wheat\_ton\_infl Price\_rice\_ton\_infl Price\_corn\_ton\_infl

1 2000 153. 313. 137

**Ποιες ήταν οι τιμές για ρύζι και καλαμπόκι την χρονιά που το σιτάρι είχε την υψηλότερη τιμή για τα τελευταία 30 χρονιά**

yearly\_price %>% select(Year, Price\_wheat\_ton\_infl, Price\_rice\_ton\_infl,Price\_corn\_ton\_infl) %>% slice\_max(yearly\_price$Price\_wheat\_ton\_infl)

Year Price\_wheat\_ton\_infl Price\_rice\_ton\_infl Price\_corn\_ton\_infl

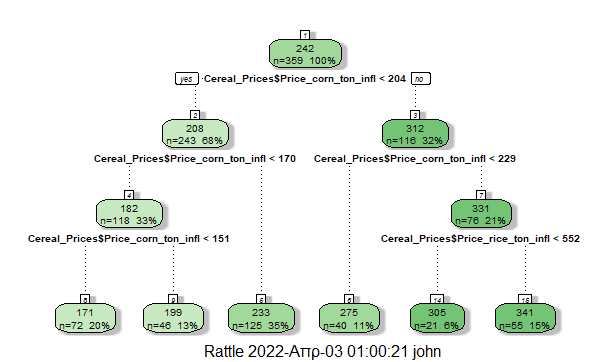
1 2012 342. 652. 346.

**Δημιουργία προβλεπτικού μοντέλου μέσω Δένδρου αποφάσεων**

tree\_model3 <- rpart(Cereal\_Prices$Price\_wheat\_ton\_infl ~ Cereal\_Prices$Price\_rice\_ton\_infl + Cereal\_Prices$Price\_corn\_ton\_infl, data = Cereal\_Prices, method = "anova")

Ο παραπάνω τύπος δημιουργεί ένα δένδρο αποφάσεων για τις τιμές του σιταριού όταν πληρούνται οι συνθήκες ,που θα δείτε στο διάγραμμα , για τις τιμές του ρυζιού και του καλαμποκιού.

fancyRpartPlot(tree\_model3)



Οι αριστερές διακλαδώσεις περιέχουν κόμβους όπου η προηγηθήσα συνθήκη ισχύει και δεξιά όπου οι συνθήκες δεν ισχύουν. Μέσα στους κόμβους παρατηρούμε την μέση τιμή του σιταριού, όπου n το πλήθος των παρατηρήσεων που σχηματίζουν το παρών κόμβο και το ποσοστό αυτών στο σύνολο των δεδομένων.

Βλέπουμε για παράδειγμα τον κόμβο 6 οπού έχει δημιουργηθεί μετά από έλεγχο 2 συνθηκών. 1η συνθήκη η τιμή του καλαμποκιού είναι μικρότερη από 204 και η 2η συνθήκη ότι η τιμή του καλαμποκιού είναι μεγαλύτερη από 170 , βλέπουμε λοιπόν ότι για τιμή καλαμποκιού 170-204 η τιμή του σιταριού θα κυμανθεί κατά μέσο όρο 233.

**Παρακάτω θα παρατεθούν τα τεταρτημόρια των ετήσιων τιμών για κάθε είδος Δημητριακού και διαγράμματα για την οπτικοποιηση τους**

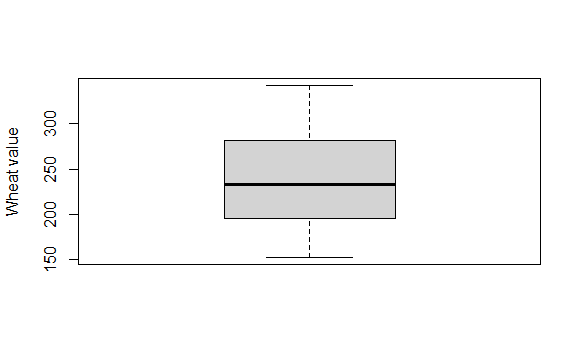
**Για το σιτάρι**

summary(yearly\_price$Price\_wheat\_ton\_infl)

Min. 1st Qu. Median Mean 3rd Qu. Max.

152.8 196.7 232.9 241.8 281.7 342.2

boxplot(yearly\_price$Price\_wheat\_ton\_infl, ylab = "Wheat value")



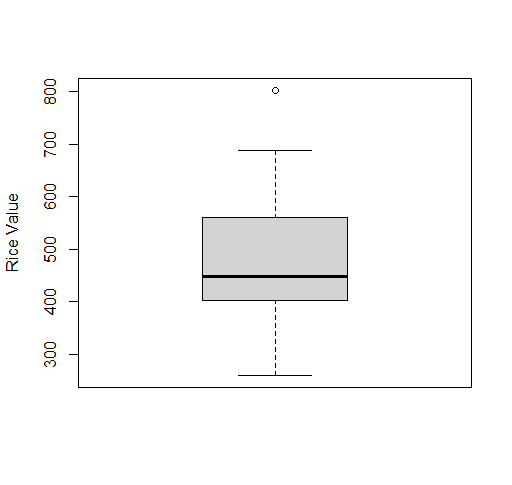
**Για το ρύζι**

summary(yearly\_price$Price\_rice\_ton\_infl)

Min. 1st Qu. Median Mean 3rd Qu. Max.

259.7 406.4 449.0 474.1 547.9 803.4

boxplot(yearly\_price$Price\_rice\_ton\_infl,ylab="Rice Value")



Από το διάγραμμα μπορούμε να διαπιστώσουμε ότι η υψηλότερη τιμή του ρυζιού 803,4 έχει μεγάλη απόκλιση από την μέση τιμή

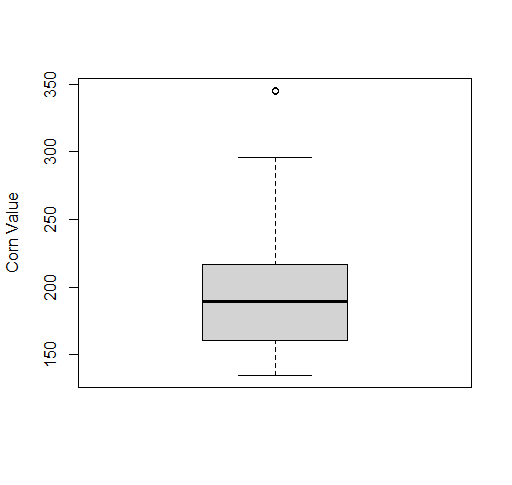
**Για το καλαμπόκι**

summary(yearly\_price$Price\_corn\_ton\_infl)

Min. 1st Qu. Median Mean 3rd Qu. Max.

134.4 162.2 189.3 201.2 216.5 345.8

boxplot(yearly\_price$Price\_corn\_ton\_infl,ylab="Corn Value")



Από το διάγραμμα μπορούμε να διαπιστώσουμε ότι η υψηλότερη τιμή του καλαμποκιού 345,8 έχει μεγάλη απόκλιση από την μέση τιμή. Επίσης είναι εμφανείς , από το πάνω όριο του θηκογραμματος ότι η τιμή του καλαμποκιού αυξάνεται

Με την παρούσα εργασία είδαμε πως κυμάνθηκαν οι τιμές των δημητριακών σε ετήσια βάση για τα τελευταία 30 χρόνια, δημιουργήθηκαν μοντέλα παλινδρόμησης και δένδρο αποφάσεων για την προβλέψει τιμών. Επίσης είδαμε την διασπορά παρατηρήσεων με την βοήθεια των boxplot για τις τιμές των δημητριακών. Όλες οι αναλύσεις έγιναν με βάσει τις σύγχρονες τιμές τους δηλαδή με βάσει τα πεδία που περιέχουν τον πληθωρισμό