

ISLR Chapter1

Intro

ACO

2025-01-02

Contents

LOAD LIB	2
WAGES DATA	2
STOCK MARKET DATA	2
FIGURE 1.3	6
Gene Expression Data	6

LOAD LIB

```
1 library(ISLR)
```

WAGES DATA

```
1 data("Wage")
2 ##### FIGURE 1.1 #####
3 ## WAGES DATA
4 par(mfrow = c(1,3))
5 # plot 2 wage ~ age
6 plot(
7   wage ~ age, data = Wage)
8 #add fitted regression line to scatterplot
9
10 # plot 2 wage ~ year
11 plot(
12   wage ~ year, data = Wage)
13 abline(lm(wage ~ year, data = Wage), col = 'blue')
14
15 # plot 2 wage ~ education
16 # boxplot(
17 #   wage ~ education, data = Wage)
18 #
19 wage_edu_box <- boxplot(
20   wage ~ education, data = Wage,
21   xaxt = "n", border = "white",
22   col = c("blue","orange","green", "black", "lightblue"),
23   boxwex = 0.3, medlwd = 1, whiskcol = "black",
24   staplecol = "black", outcol = "red", cex = 0.3, outpch=19,
25   main = "Wages by Education Levels")
26 axis(
27   side = 1, at = 1:length(wage_edu_box$names),
28   labels = paste(
29     # sub removes first occurrence of space and gsub removes all
30     # https://stackoverflow.com/questions/58196481/substitute-up-until-first-dash-with-regex?noredirect
31     sub(pattern = ".*? ", x = wage_edu_box$names,
32       replacement = ""), "\n(n=", wage_edu_box$n, ")", sep=""),
33   mgp = c(3,2,0))
```

```
1 #
2 par(mfrow = c(1,1))
3 #
```

STOCK MARKET DATA

```
1 # FIGURE 1.2 #####
2 # STOCK MARKET DATA
```

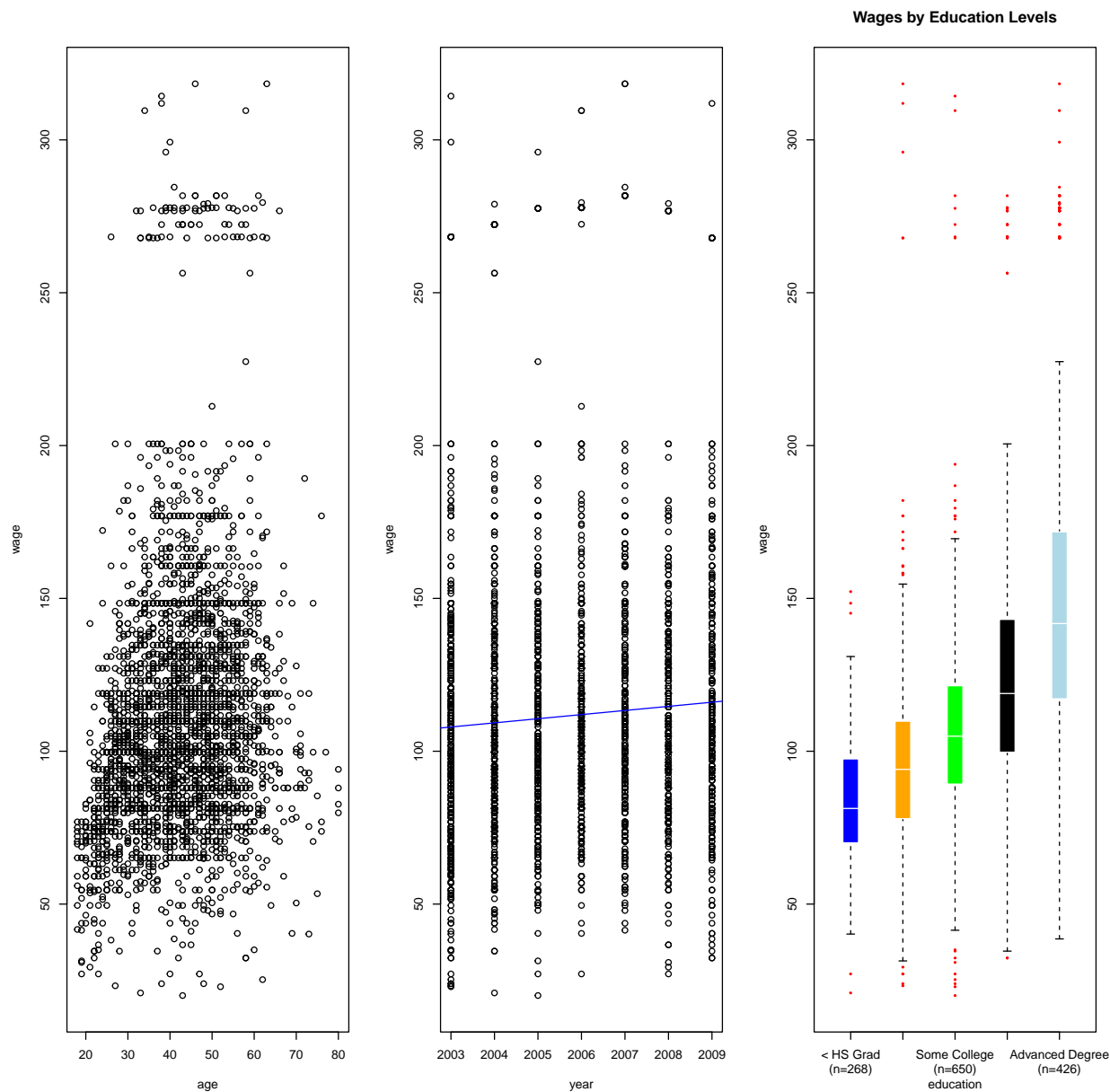


Figure 1: FIGURE 1.1

```

3 data("Smarket")
4 #
5 head(Smarket)

```

```

1 ##   Year   Lag1   Lag2   Lag3   Lag4   Lag5 Volume  Today Direction
2 ## 1 2001  0.381 -0.192 -2.624 -1.055  5.010 1.1913  0.959      Up
3 ## 2 2001  0.959  0.381 -0.192 -2.624 -1.055 1.2965  1.032      Up
4 ## 3 2001  1.032  0.959  0.381 -0.192 -2.624 1.4112 -0.623     Down
5 ## 4 2001 -0.623  1.032  0.959  0.381 -0.192 1.2760  0.614      Up
6 ## 5 2001  0.614 -0.623  1.032  0.959  0.381 1.2057  0.213      Up
7 ## 6 2001  0.213  0.614 -0.623  1.032  0.959 1.3491  1.392      Up

```

```

1 #
2 par(mfrow = c(1,3))
3 lag1_direction_box <- boxplot(
4   Lag1 ~ Direction, data = Smarket,
5   xaxt = "n", border = "white",
6   col = c("blue","orange"),
7   boxwex = 0.3, medlwd = 1, whiskcol = "black",
8   staplecol = "black", outcol = "red", cex = 0.3, outpch=19,
9   main = "Yesterday", ylab = "Percentage Change in S&P",
10  xlab = "Today's Direction")
11 axis(
12   side = 1, at = 1:length(lag1_direction_box$names),
13   labels = paste(
14     lag1_direction_box$names, "\n(n=", lag1_direction_box$n, ")", sep=""),
15   mgp = c(3,2,0))
16 #
17 lag2_direction_box <- boxplot(
18   Lag2 ~ Direction, data = Smarket,
19   xaxt = "n", border = "white",
20   col = c("blue","orange"),
21   boxwex = 0.3, medlwd = 1, whiskcol = "black",
22   staplecol = "black", outcol = "red", cex = 0.3, outpch=19,
23   main = "Two Days Previous", ylab = "Percentage Change in S&P",
24   xlab = "Today's Direction")
25 axis(
26   side = 1, at = 1:length(lag2_direction_box$names),
27   labels = paste(
28     lag2_direction_box$names, "\n(n=", lag2_direction_box$n, ")", sep=""),
29   mgp = c(3,2,0))
30 #
31 lag3_direction_box <- boxplot(
32   Lag3 ~ Direction, data = Smarket,
33   xaxt = "n", border = "white",
34   col = c("blue","orange"),
35   boxwex = 0.3, medlwd = 1, whiskcol = "black",
36   staplecol = "black", outcol = "red", cex = 0.3, outpch=19,
37   main = "Three Days Previous", ylab = "Percentage Change in S&P",
38   xlab = "Today's Direction")
39 axis(
40   side = 1, at = 1:length(lag3_direction_box$names),
41   labels = paste(

```

```

42 lag3_direction_box$names, "\n(n=", lag3_direction_box$n, ")", sep=""),
43 mgp = c(3,2,0))

```

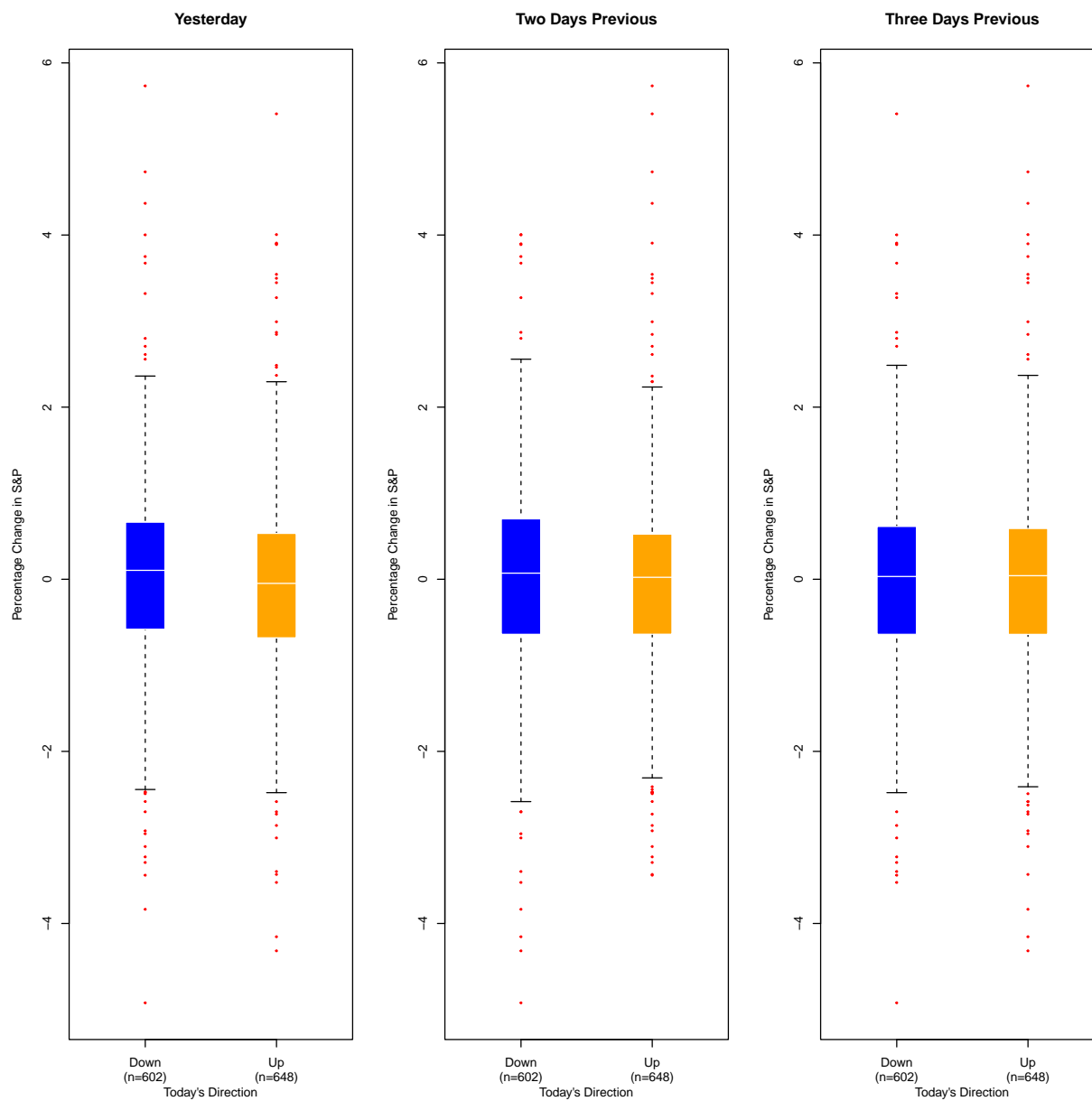


Figure 2: FIGURE 1.2

```

1 #
2 par(mfrow = c(1,1))
3 #

```

FIGURE 1.3

```
1  # FIGURE 1.3 #####
2  #
3  model1 = MASS::qda(
4    Direction~., data = subset(Smarket, Year < 2005))
5  prediction_prob = predict(
6    model1, newdata = subset(Smarket, Year >= 2005),
7    type = 'response')$posterior
8  # prediction_prob
9  #
10 boxplot(
11   prediction_prob,
12   ylab = "Predicted Probability",
13   xlab = "Today's Direction"
14 )
```

```
1  #
```

Gene Expression Data

```
1  ## Gene Expression Data
2
3  data("NCI60")
4  #
5  # head(NCI60)
6  #
7  nc160_pca = prcomp(NCI60$data, scale. = T, center = T)
```

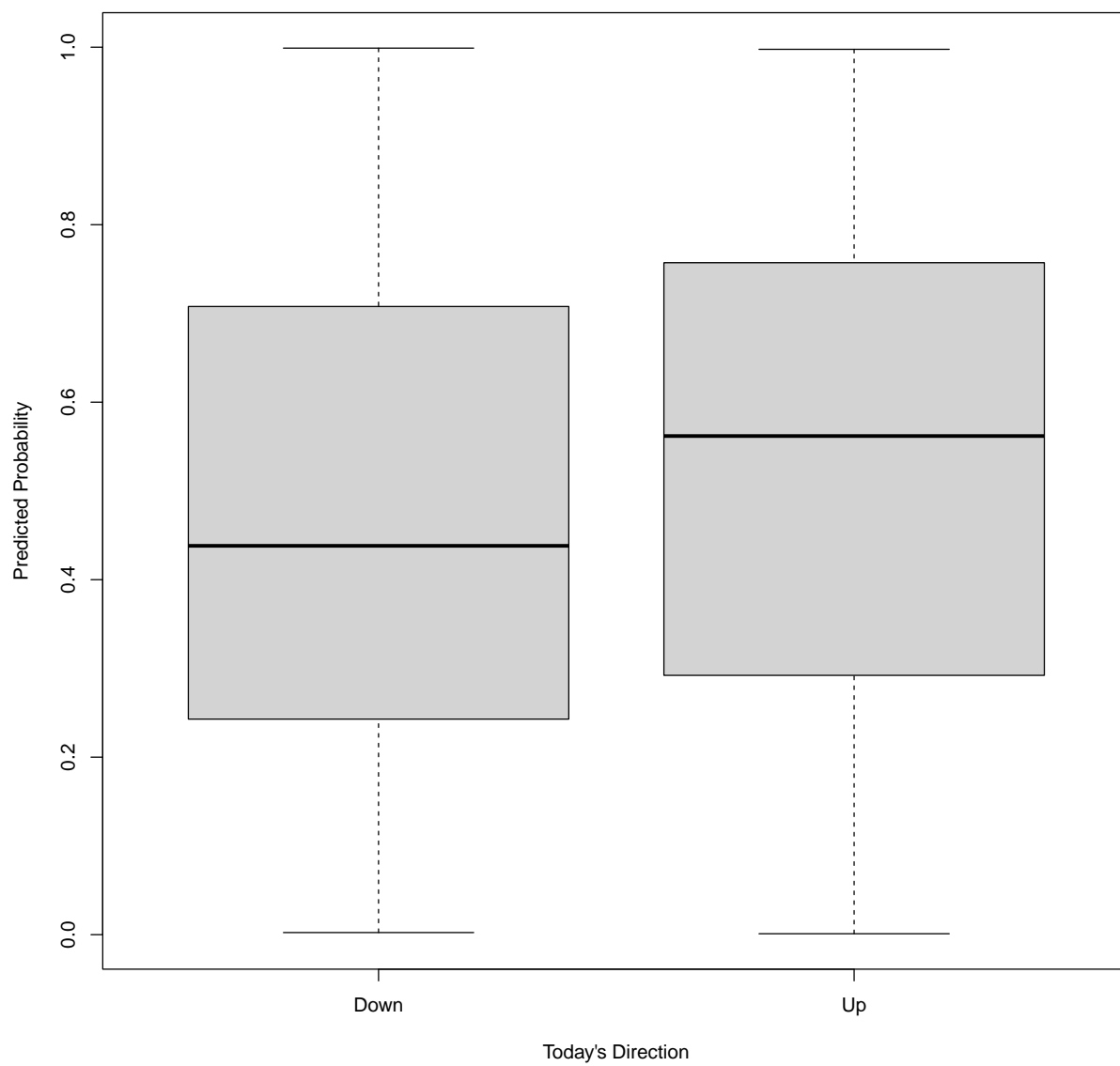


Figure 3: FIGURE 1.3