

Overview of “Weekly” Dataset from “ISLR” Package

Julian Hatwell

February 8, 2016

This document provides a brief overview of the Weekly dataset in the ISLR R package.

```
##      Year          Lag1          Lag2          Lag3
## Min. :1990  Min. :-18.1950  Min. :-18.1950  Min. :-18.1950
## 1st Qu.:1995  1st Qu.: -1.1540  1st Qu.: -1.1540  1st Qu.: -1.1580
## Median :2000  Median :  0.2410  Median :  0.2410  Median :  0.2410
## Mean   :2000  Mean   :  0.1506  Mean   :  0.1511  Mean   :  0.1472
## 3rd Qu.:2005  3rd Qu.:  1.4050  3rd Qu.:  1.4090  3rd Qu.:  1.4090
## Max.   :2010  Max.   : 12.0260  Max.   : 12.0260  Max.   : 12.0260
##          Lag4          Lag5          Volume
## Min. :-18.1950  Min. :-18.1950  Min. : 0.08747
## 1st Qu.: -1.1580  1st Qu.: -1.1660  1st Qu.: 0.33202
## Median :  0.2380  Median :  0.2340  Median : 1.00268
## Mean   :  0.1458  Mean   :  0.1399  Mean   : 1.57462
## 3rd Qu.:  1.4090  3rd Qu.:  1.4050  3rd Qu.: 2.05373
## Max.   : 12.0260  Max.   : 12.0260  Max.   : 9.32821
##          Today          Direction
## Min. :-18.1950  Down:484
## 1st Qu.: -1.1540  Up  :605
## Median :  0.2410
## Mean   :  0.1499
## 3rd Qu.:  1.4050
## Max.   : 12.0260
```

From the summary, and the associated help (not shown), the following observations can be made:

The dataframe contains 1089 rows and 9 columns.

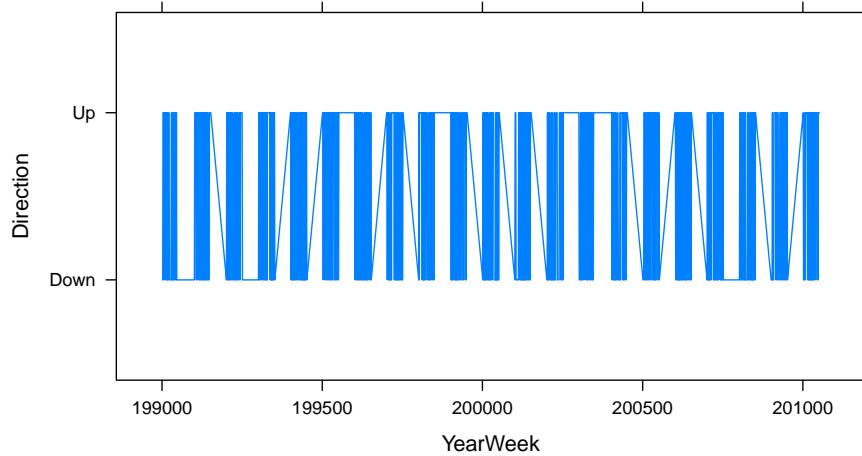


Figure 1: Time Series plot of Direction against YearWeek

```

## 
## Call:
## glm(formula = fmla1, family = binomial, data = df)
## 
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max 
## -1.306  -1.267   1.060   1.087   1.115 
## 
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept) 15.045416  20.235269   0.744   0.457    
## Year        -0.007411   0.010117  -0.733   0.464    
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1495.7 on 1087 degrees of freedom
## AIC: 1499.7
## 
## Number of Fisher Scoring iterations: 3
## 
## Call:
## glm(formula = fmla1, family = binomial, data = df)
## 
```

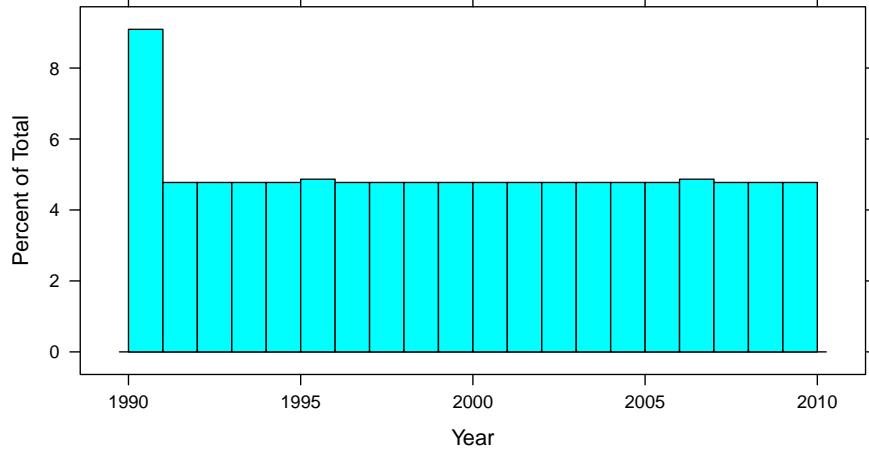


Figure 2: Histogram of the Year variable

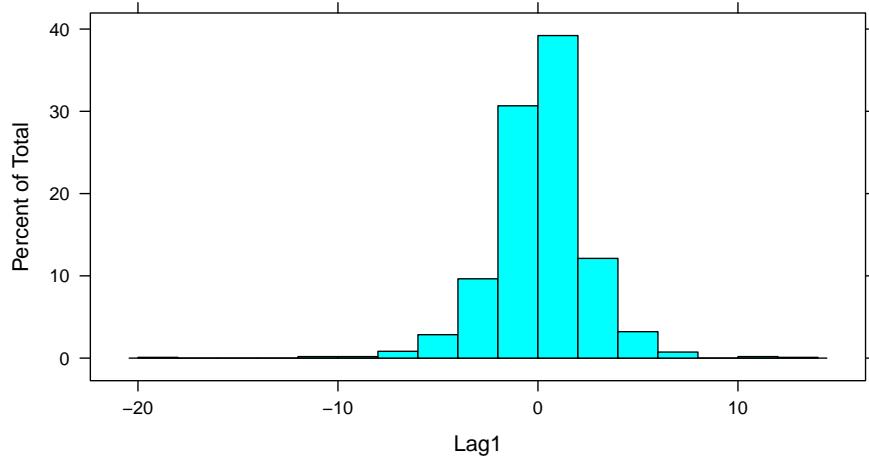


Figure 3: Histogram of the Lag1 variable

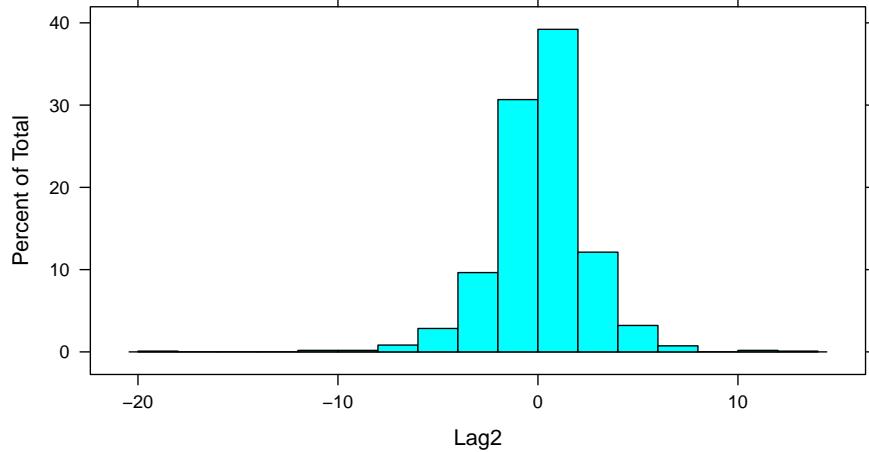


Figure 4: Histogram of the Lag2 variable

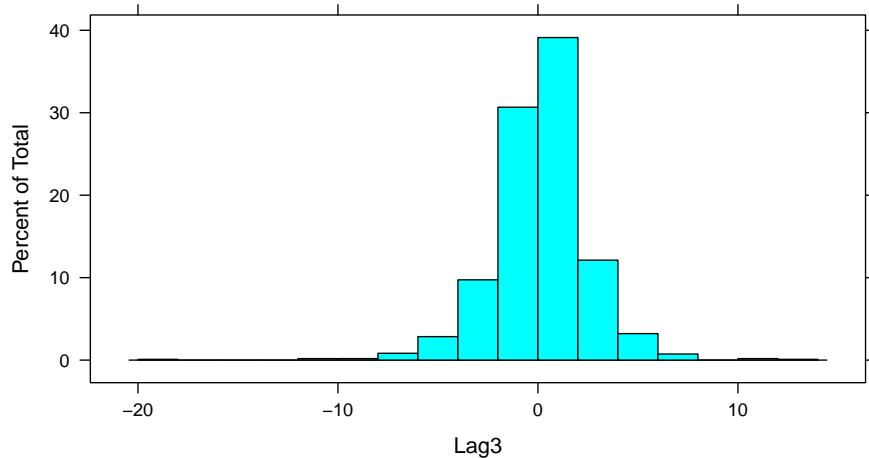


Figure 5: Histogram of the Lag3 variable

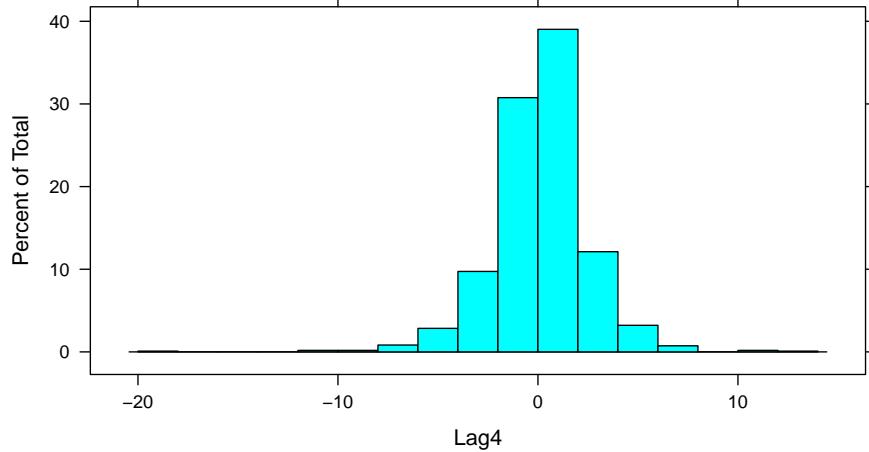


Figure 6: Histogram of the Lag4 variable

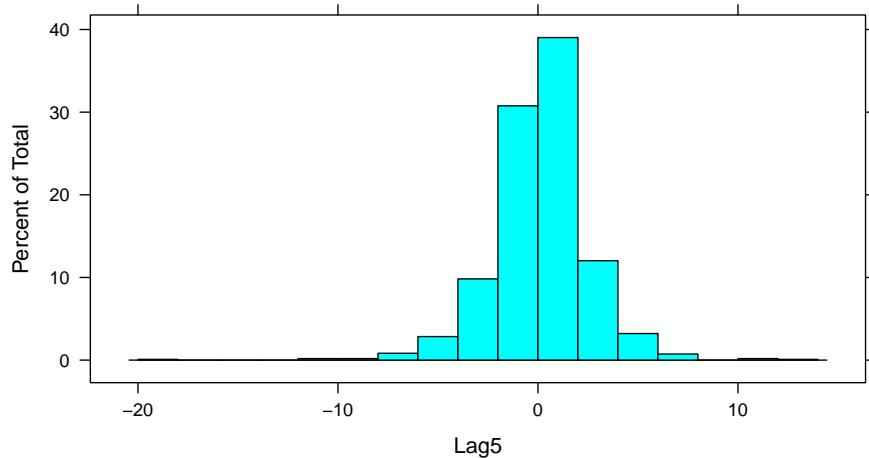


Figure 7: Histogram of the Lag5 variable

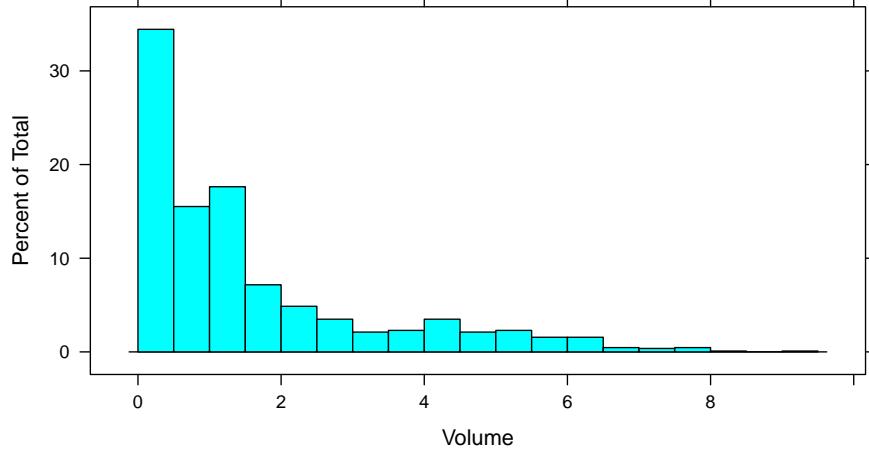


Figure 8: Histogram of the Volume variable

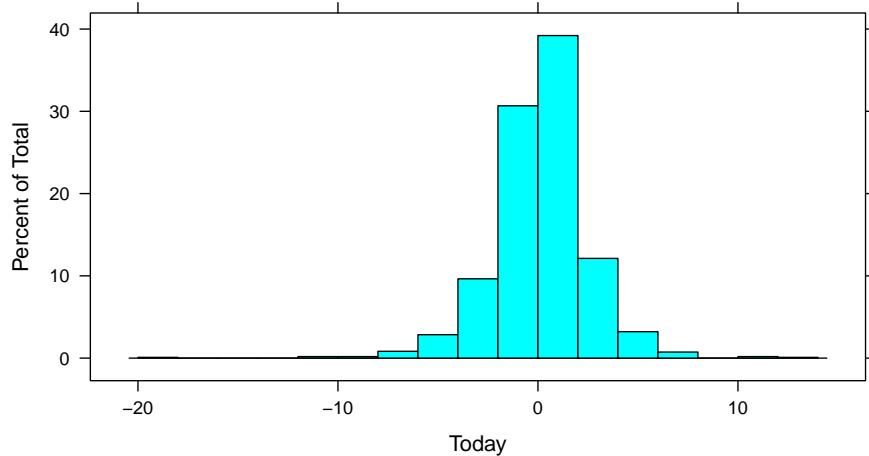


Figure 9: Histogram of the Today variable

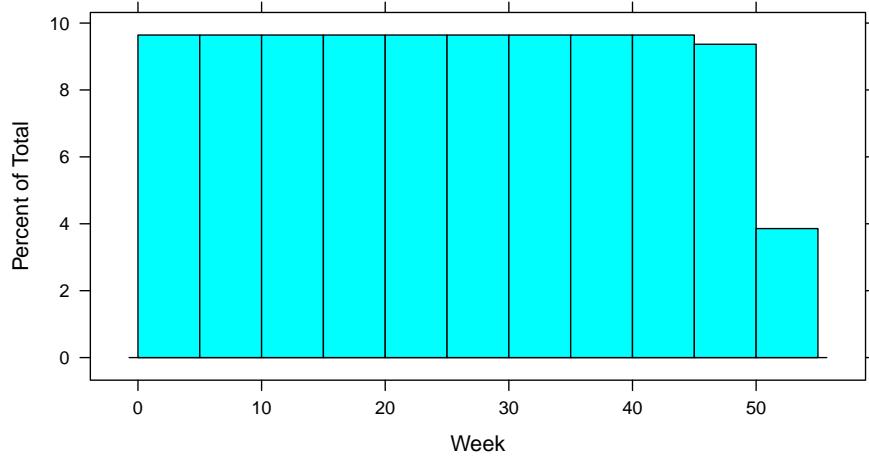


Figure 10: Histogram of the Week variable

```

## Deviance Residuals:
##      Min      1Q  Median      3Q     Max 
## -1.456  -1.263   1.041   1.087   1.277 
## 
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)    
## (Intercept) 0.23024   0.06124   3.760  0.00017 ***
## Lag1        -0.04313   0.02622  -1.645  0.10001    
## ---        
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1493.5 on 1087 degrees of freedom
## AIC: 1497.5
## 
## Number of Fisher Scoring iterations: 4
## 
## 
## Call:
## glm(formula = fmla1, family = binomial, data = df)
## 
## Deviance Residuals:
##      Min      1Q  Median      3Q     Max 
## -1.456  -1.263   1.041   1.087   1.277 
## 
```

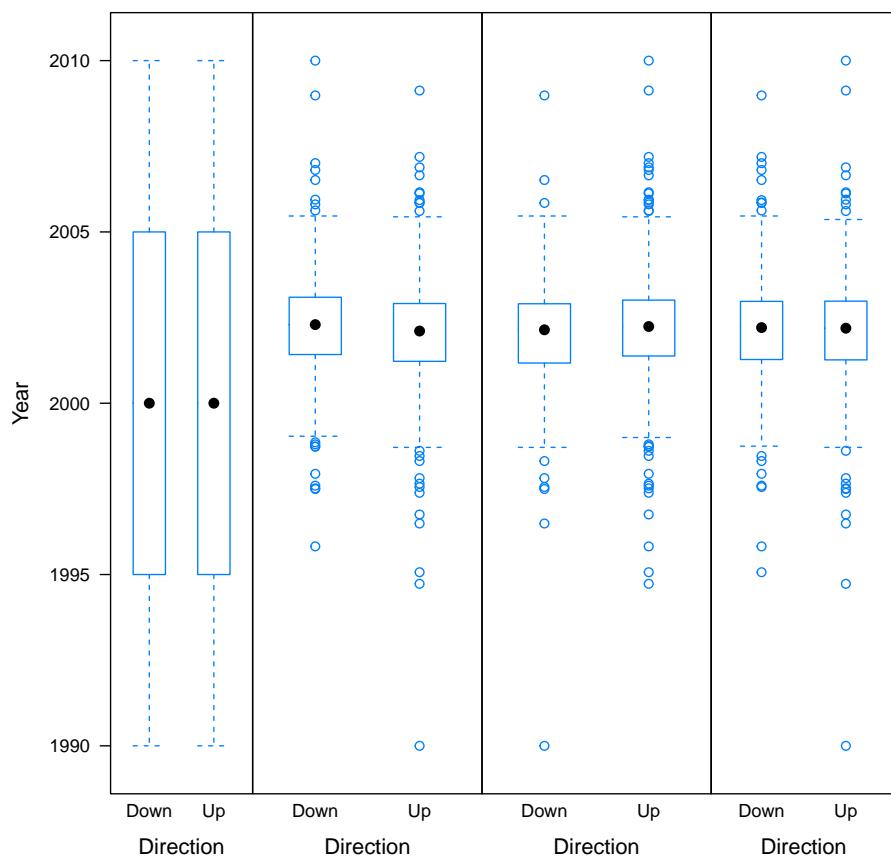


Figure 11: Boxplot of the dependent variable Direction by each factor variable

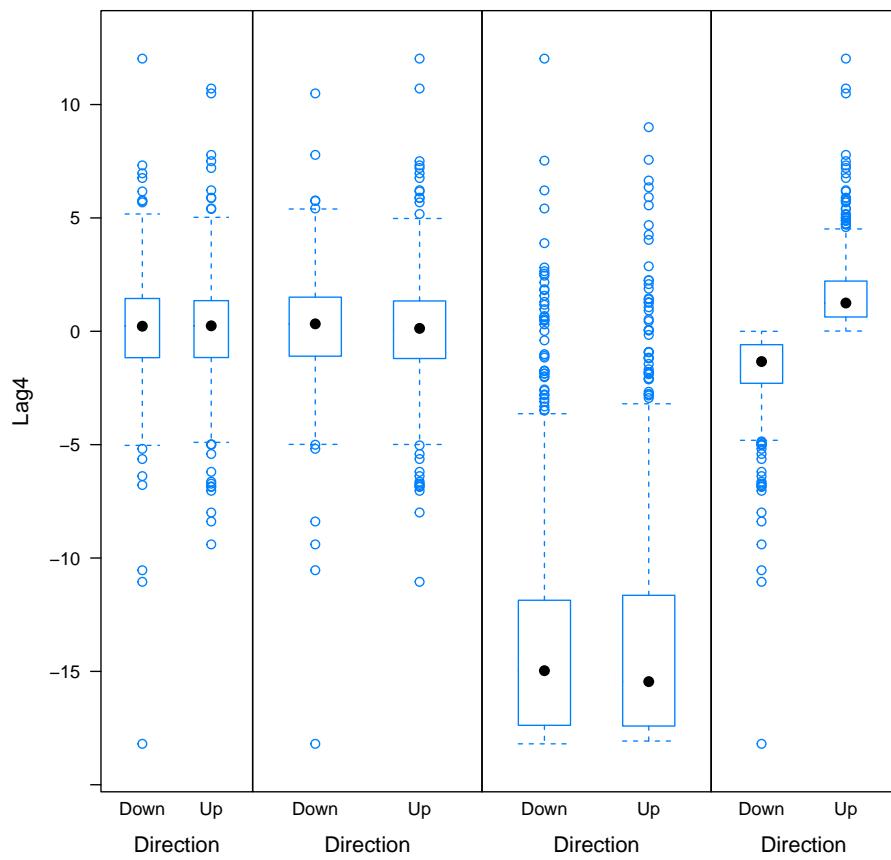


Figure 12: Boxplot of the dependent variable Direction by each factor variable

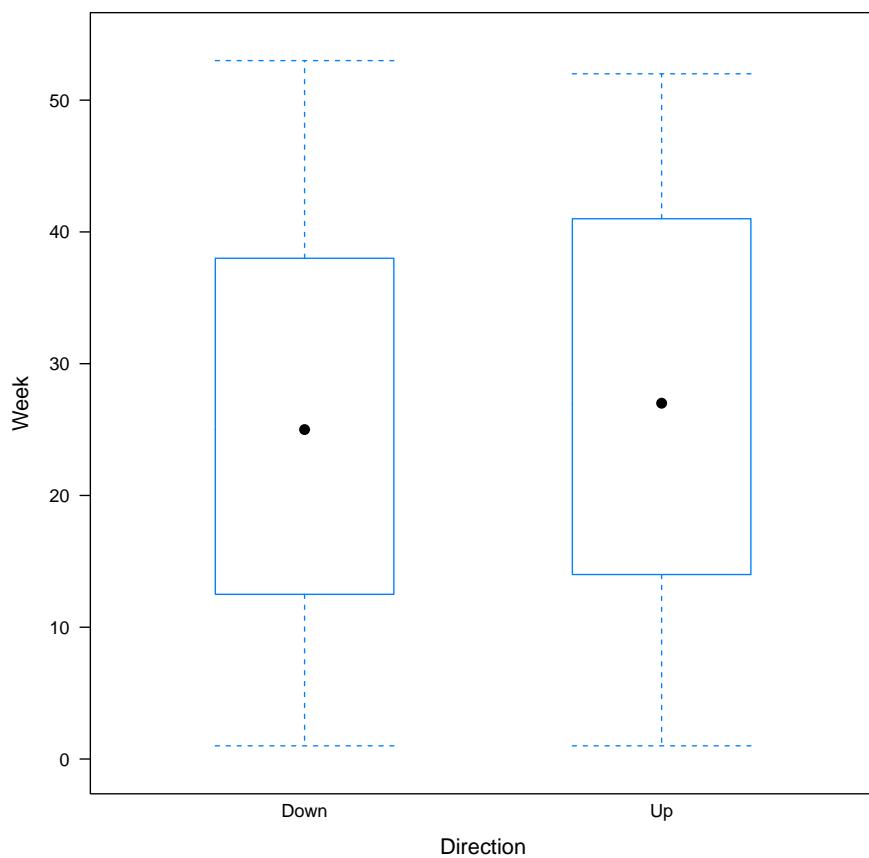


Figure 13: Boxplot of the dependent variable Direction by each factor variable

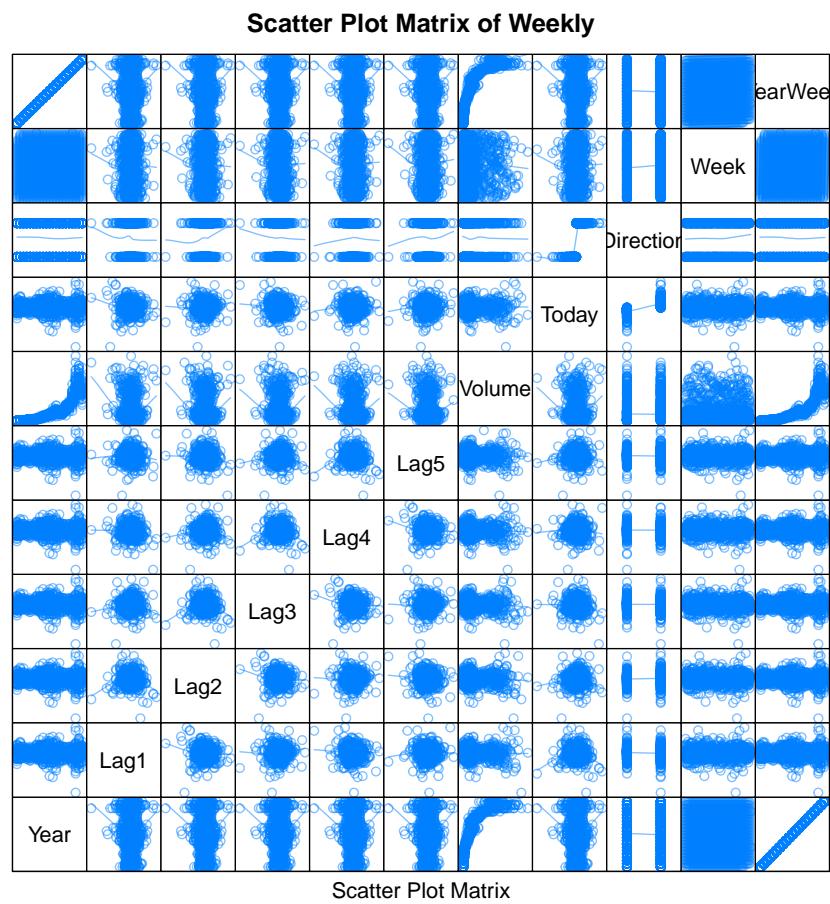


Figure 14: multi-variate comparisons

Correlogram of Weekly

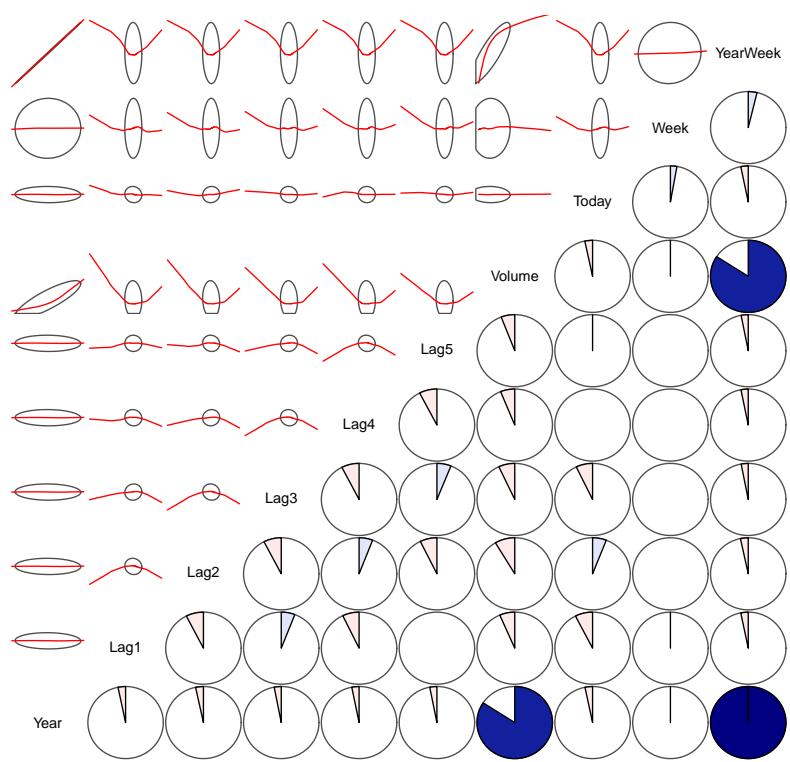


Figure 15: Correlogram

```

## -1.564 -1.267  1.008  1.086  1.386
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.21473   0.06123  3.507 0.000453 ***
## Lag2        0.06279   0.02636  2.382 0.017230 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1490.4 on 1087 degrees of freedom
## AIC: 1494.4
##
## Number of Fisher Scoring iterations: 4
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.366 -1.270  1.063  1.087  1.181
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.22615   0.06114  3.699 0.000217 ***
## Lag3        -0.01960   0.02593 -0.756 0.449892
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1495.6 on 1087 degrees of freedom
## AIC: 1499.6
##
## Number of Fisher Scoring iterations: 3
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df)
##
## Deviance Residuals:
```

```

##      Min      1Q Median      3Q      Max
## -1.416  -1.270   1.065   1.086   1.161
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.22580   0.06113  3.694 0.000221 ***
## Lag4        -0.01757   0.02592 -0.678 0.497930
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1495.7 on 1087 degrees of freedom
## AIC: 1499.7
##
## Number of Fisher Scoring iterations: 3
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df)
##
## Deviance Residuals:
##      Min      1Q Median      3Q      Max
## -1.399  -1.270   1.068   1.086   1.161
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.22539   0.06112  3.688 0.000226 ***
## Lag5        -0.01552   0.02590 -0.599 0.548980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1495.8 on 1087 degrees of freedom
## AIC: 1499.8
##
## Number of Fisher Scoring iterations: 3
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df)
##

```

```

## Deviance Residuals:
##      Min      1Q  Median      3Q     Max
## -1.287 -1.277  1.072  1.080  1.145
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.25694   0.08350  3.077 0.00209 **
## Volume     -0.02142   0.03609 -0.594 0.55273
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1495.9 on 1087 degrees of freedom
## AIC: 1499.9
##
## Number of Fisher Scoring iterations: 3
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df)
##
## Deviance Residuals:
##      Min      1Q  Median      3Q     Max
## -0.004844  0.000000  0.000000  0.000000  0.004870
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.57     150.50 -0.050    0.960
## Today       1891.30  20807.77  0.091    0.928
## 
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1.4962e+03 on 1088 degrees of freedom
## Residual deviance: 4.7189e-05 on 1087 degrees of freedom
## AIC: 4
##
## Number of Fisher Scoring iterations: 25
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df)
##
## Deviance Residuals:

```

```

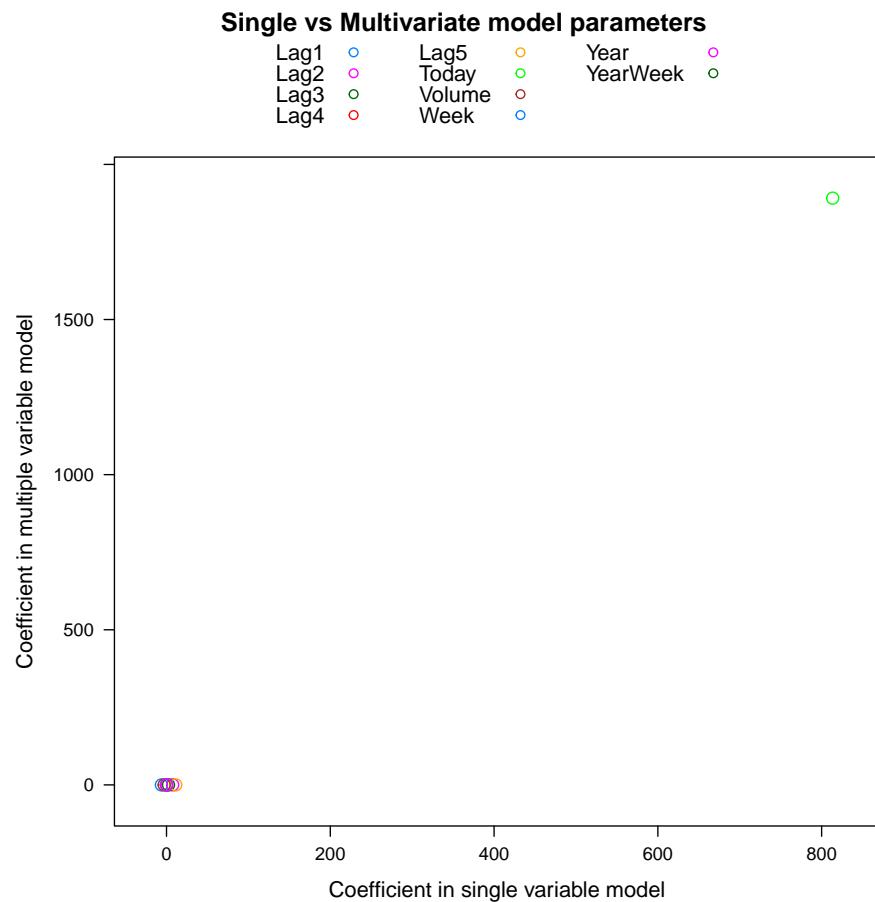
##      Min      1Q Median      3Q      Max
## -1.354 -1.263  1.024  1.091  1.157
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 0.042046  0.123329  0.341  0.7332
## Week       0.006871  0.004079  1.685  0.0921 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1493.4 on 1087 degrees of freedom
## AIC: 1497.4
##
## Number of Fisher Scoring iterations: 4
##
##
## Call:
## glm(formula = fmla1, family = binomial, data = df)
##
## Deviance Residuals:
##      Min      1Q Median      3Q      Max
## -1.305 -1.268  1.061  1.087  1.114
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) 1.418e+01  2.022e+01  0.701  0.483
## YearWeek   -6.978e-05  1.011e-04 -0.690  0.490
## ---
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1496.2 on 1088 degrees of freedom
## Residual deviance: 1495.7 on 1087 degrees of freedom
## AIC: 1499.7
##
## Number of Fisher Scoring iterations: 3
##
## Call:
## glm(formula = fmla, family = binomial, data = df)
##
## Deviance Residuals:
##      Min      1Q Median      3Q      Max
## -1.962e-03 -2.000e-08  2.000e-08  2.000e-08  1.609e-03

```

```

##
## Coefficients: (1 not defined because of singularities)
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept) -9.489e+02  4.472e+05 -0.002   0.998
## Year        4.752e-01  2.250e+02  0.002   0.998
## Lag1       -6.222e+00  1.807e+03 -0.003   0.997
## Lag2        7.588e+00  1.103e+03  0.007   0.995
## Lag3        2.462e+00  1.603e+03  0.002   0.999
## Lag4       -1.966e-01  3.878e+02 -0.001   1.000
## Lag5        1.130e+01  6.988e+02  0.016   0.987
## Volume     -2.872e+00  1.607e+03 -0.002   0.999
## Today       8.135e+02  1.648e+04  0.049   0.961
## Week       -1.041e-01  8.353e+01 -0.001   0.999
## YearWeek      NA        NA        NA        NA
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 1.4962e+03 on 1088 degrees of freedom
## Residual deviance: 1.0658e-05 on 1079 degrees of freedom
## AIC: 20
##
## Number of Fisher Scoring iterations: 25

```



```

Week <- numeric(0)
for (yr in unique(df$Year)) {
  Week <- c(Week, 1:table(Weekly$Year)[as.character(yr)])
}
df <- mutate(df, Week = Week, YearWeek = ts(100 *
  Year + Week))

```