W271 Section 3 Lab 3

Kiersten Henderson, Zhaoning Yu, Daghan Altas

11/17/2017

Task 5:

1. Read AMAZ.csv and UMCSENT.csv into R as R DataFrames

```
AMAZ_df <- read.csv("./AMAZ.csv")
UMCSENT_df <- read.csv("./UMCSENT.csv")

rbind(head(AMAZ_df), tail(AMAZ_df))
```

```
##           Index AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume
## 1    2007-01-03     20.00     20.00    16.00      16.00         650
## 2    2007-01-04     20.00     20.00    20.00      20.00          67
## 3    2007-01-08     19.20     22.00    19.20      22.00        1801
## 4    2007-01-09     22.00     22.00    20.80      20.80         356
## 5    2007-01-10     20.80     20.80    20.80      20.80         438
## 6    2007-01-11     20.80     21.60    20.80      21.60        2318
## 1174 2013-01-04      0.88      0.88     0.80       0.80        3850
## 1175 2013-01-07      0.80      1.00     0.80       1.00        2715
## 1176 2013-01-08      0.80      0.80     0.68       0.68        4668
## 1177 2013-01-09      0.88      0.88     0.80       0.80        2750
## 1178 2013-01-11      0.80      0.80     0.80       0.80        3000
## 1179 2013-01-15      0.68      0.68     0.68       0.68        1000
```

```
dim(AMAZ_df)
```

```
## [1] 1179    6
```

```
str(AMAZ_df)
```

```
## 'data.frame':    1179 obs. of  6 variables:
##  $ Index      : Factor w/ 1179 levels "2007-01-03","2007-01-04",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ AMAZ.Open  : num  20 20 19.2 22 20.8 20.8 22 21.6 22 23.2 ...
##  $ AMAZ.High  : num  20 20 22 22 20.8 21.6 22 21.6 22 23.2 ...
##  $ AMAZ.Low   : num  16 20 19.2 20.8 20.8 20.8 22 21.2 21.6 22.8 ...
##  $ AMAZ.Close : num  16 20 22 20.8 20.8 21.6 22 21.2 21.6 22.8 ...
##  $ AMAZ.Volume: int  650 67 1801 356 438 2318 306 925 2138 527 ...
```

```
summary(AMAZ_df)
```

```
##         Index        AMAZ.Open       AMAZ.High        AMAZ.Low
##  2007-01-03:   1   Min.   : 0.16   Min.   : 0.200   Min.   : 0.080
##  2007-01-04:   1   1st Qu.: 0.80   1st Qu.: 0.800   1st Qu.: 0.720
##  2007-01-08:   1   Median : 1.08   Median : 1.120   Median : 1.000
##  2007-01-09:   1   Mean   : 4.83   Mean   : 4.954   Mean   : 4.696
##  2007-01-10:   1   3rd Qu.: 6.00   3rd Qu.: 6.400   3rd Qu.: 5.650
##  2007-01-11:   1   Max.   :24.40   Max.   :26.000   Max.   :24.400
##  (Other)   :1173   NA's   :259     NA's   :259      NA's   :259
##    AMAZ.Close      AMAZ.Volume
##  Min.   : 0.080   Min.   :    0
```

```
##  1st Qu.: 0.620   1st Qu.:   25
##  Median : 1.000   Median :  312
##  Mean   : 4.129   Mean   : 1499
##  3rd Qu.: 4.000   3rd Qu.: 1250
##  Max.   :25.600   Max.   :68900
##
```

```
rbind(head(UMCSENT_df, 15), tail(UMCSENT_df, 15))
```

```
##          Index UMCSENT
## 1   1978-01-01    83.7
## 2   1978-02-01    84.3
## 3   1978-03-01    78.8
## 4   1978-04-01    81.6
## 5   1978-05-01    82.9
## 6   1978-06-01    80.0
## 7   1978-07-01    82.4
## 8   1978-08-01    78.4
## 9   1978-09-01    80.4
## 10  1978-10-01    79.3
## 11  1978-11-01    75.0
## 12  1978-12-01    66.1
## 13  1979-01-01    72.1
## 14  1979-02-01    73.9
## 15  1979-03-01    68.4
## 463 2016-07-01    90.0
## 464 2016-08-01    89.8
## 465 2016-09-01    91.2
## 466 2016-10-01    87.2
## 467 2016-11-01    93.8
## 468 2016-12-01    98.2
## 469 2017-01-01    98.5
## 470 2017-02-01    96.3
## 471 2017-03-01    96.9
## 472 2017-04-01    97.0
## 473 2017-05-01    97.1
## 474 2017-06-01    95.1
## 475 2017-07-01    93.4
## 476 2017-08-01    96.8
## 477 2017-09-01    95.1
```

```
dim(UMCSENT_df)
```

```
## [1] 477   2
```

```
str(UMCSENT_df)
```

```
## 'data.frame':    477 obs. of  2 variables:
##  $ Index  : Factor w/ 477 levels "1978-01-01","1978-02-01",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ UMCSENT: num  83.7 84.3 78.8 81.6 82.9 80 82.4 78.4 80.4 79.3 ...
```

```
summary(UMCSENT_df)
```

```
##         Index        UMCSENT
##  1978-01-01:  1   Min.   : 51.70
##  1978-02-01:  1   1st Qu.: 76.10
##  1978-03-01:  1   Median : 89.30
```

```
##  1978-04-01:  1   Mean   : 85.69
##  1978-05-01:  1   3rd Qu.: 94.30
##  1978-06-01:  1   Max.   :112.00
##  (Other)   :471
```

2. Convert them to xts objects

```
library(xts)

AMAZ <- as.xts(AMAZ_df[, -1], order.by = as.POSIXct(AMAZ_df$Index,
    format = "%Y-%m-%d"))
UMCSENT <- as.xts(UMCSENT_df[, -1], order.by = as.POSIXct(UMCSENT_df$Index,
    format = "%Y-%m-%d"))

rbind(head(AMAZ), tail(AMAZ))
```

```
##            AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume
## 2007-01-03     20.00     20.00    16.00      16.00         650
## 2007-01-04     20.00     20.00    20.00      20.00          67
## 2007-01-08     19.20     22.00    19.20      22.00        1801
## 2007-01-09     22.00     22.00    20.80      20.80         356
## 2007-01-10     20.80     20.80    20.80      20.80         438
## 2007-01-11     20.80     21.60    20.80      21.60        2318
## 2013-01-04      0.88      0.88     0.80       0.80        3850
## 2013-01-07      0.80      1.00     0.80       1.00        2715
## 2013-01-08      0.80      0.80     0.68       0.68        4668
## 2013-01-09      0.88      0.88     0.80       0.80        2750
## 2013-01-11      0.80      0.80     0.80       0.80        3000
## 2013-01-15      0.68      0.68     0.68       0.68        1000
```

```
rbind(head(UMCSENT), tail(UMCSENT))
```

```
##            [,1]
## 1978-01-01 83.7
## 1978-02-01 84.3
## 1978-03-01 78.8
## 1978-04-01 81.6
## 1978-05-01 82.9
## 1978-06-01 80.0
## 2017-04-01 97.0
## 2017-05-01 97.1
## 2017-06-01 95.1
## 2017-07-01 93.4
## 2017-08-01 96.8
## 2017-09-01 95.1
```

3. Merge the two set of series together, perserving all of the obserbvations in both set of series

a. fill all of the missing values of the UMCSENT series with -9999

```
AMAZ_UMCSENT_outer <- merge(AMAZ, UMCSENT, join = "outer", fill = -9999)

# in order to be sure that the merge and subsequent
# manipulations occur as we intend, we monitor the head and
# tail, but also a subset of observations where there are
# observations for the AMAZ series but not the UMCSENT
# series.
```

```
rbind(head(AMAZ_UMCSENT_outer), AMAZ_UMCSENT_outer["2007-01-26/2007-02-09"],
    tail(AMAZ_UMCSENT_outer))
```

```
##            AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume UMCSENT
## 1978-01-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    83.7
## 1978-02-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    84.3
## 1978-03-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    78.8
## 1978-04-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    81.6
## 1978-05-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    82.9
## 1978-06-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    80.0
## 2007-01-26      22.0      22.0     22.0       22.0         250 -9999.0
## 2007-01-29      22.8      23.2     22.0       23.2         986 -9999.0
## 2007-01-31      23.6      24.0     23.6       24.0         125 -9999.0
## 2007-02-01      24.0      24.0     24.0       24.0         270    91.3
## 2007-02-02      23.6      24.0     23.6       24.0         729 -9999.0
## 2007-02-05      24.0      25.6     24.0       25.6         375 -9999.0
## 2007-02-06      24.4      24.4     24.4       24.4         142 -9999.0
## 2007-02-09      24.0      24.0     23.2       23.6        2690 -9999.0
## 2017-04-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    97.0
## 2017-05-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    97.1
## 2017-06-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    95.1
## 2017-07-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    93.4
## 2017-08-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    96.8
## 2017-09-01   -9999.0   -9999.0  -9999.0    -9999.0       -9999    95.1
```

b. then create a new series, named UMCSENT02, from the original UMCSENT series replace all of the -9999 with NAs

```
UMCSENT02 <- xts(AMAZ_UMCSENT_outer)
UMCSENT02[UMCSENT02 <= -9999] <- NA
rbind(head(UMCSENT02), UMCSENT02["2007-01-26/2007-02-09"], tail(UMCSENT02))
```

```
##            AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume UMCSENT
## 1978-01-01        NA        NA       NA         NA           NA    83.7
## 1978-02-01        NA        NA       NA         NA           NA    84.3
## 1978-03-01        NA        NA       NA         NA           NA    78.8
## 1978-04-01        NA        NA       NA         NA           NA    81.6
## 1978-05-01        NA        NA       NA         NA           NA    82.9
## 1978-06-01        NA        NA       NA         NA           NA    80.0
## 2007-01-26      22.0      22.0     22.0       22.0          250      NA
## 2007-01-29      22.8      23.2     22.0       23.2          986      NA
## 2007-01-31      23.6      24.0     23.6       24.0          125      NA
## 2007-02-01      24.0      24.0     24.0       24.0          270    91.3
## 2007-02-02      23.6      24.0     23.6       24.0          729      NA
## 2007-02-05      24.0      25.6     24.0       25.6          375      NA
## 2007-02-06      24.4      24.4     24.4       24.4          142      NA
## 2007-02-09      24.0      24.0     23.2       23.6         2690      NA
## 2017-04-01        NA        NA       NA         NA           NA    97.0
## 2017-05-01        NA        NA       NA         NA           NA    97.1
## 2017-06-01        NA        NA       NA         NA           NA    95.1
## 2017-07-01        NA        NA       NA         NA           NA    93.4
## 2017-08-01        NA        NA       NA         NA           NA    96.8
## 2017-09-01        NA        NA       NA         NA           NA    95.1
```

c. then create a new series, named UMCSENT03, and replace the NAs with the last observation

```r
UMCSENT03 <- xts(UMCSENT02)

# depending on how we will analyze the merged series, we may
# or may not want to remove NAs that cannot be replaced with
# the last observation from the series.  First, we impute NAs
# while keeping NAs at the beginning of the series (there is
# no observation prior for these NA values)
UMCSENT03a <- na.locf(UMCSENT03, na.rm = TRUE, fromLast = FALSE)
rbind(head(UMCSENT03a), UMCSENT03a["2007-01-26/2007-02-09"],
    tail(UMCSENT03a))
```

```
##            AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume UMCSENT
## 2007-01-03     20.00     20.00    16.00      16.00         650    96.9
## 2007-01-04     20.00     20.00    20.00      20.00          67    96.9
## 2007-01-08     19.20     22.00    19.20      22.00        1801    96.9
## 2007-01-09     22.00     22.00    20.80      20.80         356    96.9
## 2007-01-10     20.80     20.80    20.80      20.80         438    96.9
## 2007-01-11     20.80     21.60    20.80      21.60        2318    96.9
## 2007-01-26     22.00     22.00    22.00      22.00         250    96.9
## 2007-01-29     22.80     23.20    22.00      23.20         986    96.9
## 2007-01-31     23.60     24.00    23.60      24.00         125    96.9
## 2007-02-01     24.00     24.00    24.00      24.00         270    91.3
## 2007-02-02     23.60     24.00    23.60      24.00         729    91.3
## 2007-02-05     24.00     25.60    24.00      25.60         375    91.3
## 2007-02-06     24.40     24.40    24.40      24.40         142    91.3
## 2007-02-09     24.00     24.00    23.20      23.60        2690    91.3
## 2017-04-01      0.68      0.68     0.68       0.68        1000    97.0
## 2017-05-01      0.68      0.68     0.68       0.68        1000    97.1
## 2017-06-01      0.68      0.68     0.68       0.68        1000    95.1
## 2017-07-01      0.68      0.68     0.68       0.68        1000    93.4
## 2017-08-01      0.68      0.68     0.68       0.68        1000    96.8
## 2017-09-01      0.68      0.68     0.68       0.68        1000    95.1
```

```r
# We also show how to impute these NA values by removing NAs
# at the beginning of the series
UMCSENT03b <- na.locf(UMCSENT03, na.rm = FALSE, fromLast = FALSE)
rbind(head(UMCSENT03b), UMCSENT03b["2007-01-26/2007-02-09"],
    tail(UMCSENT03b))
```

```
##            AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume UMCSENT
## 1978-01-01        NA        NA       NA         NA          NA    83.7
## 1978-02-01        NA        NA       NA         NA          NA    84.3
## 1978-03-01        NA        NA       NA         NA          NA    78.8
## 1978-04-01        NA        NA       NA         NA          NA    81.6
## 1978-05-01        NA        NA       NA         NA          NA    82.9
## 1978-06-01        NA        NA       NA         NA          NA    80.0
## 2007-01-26     22.00     22.00    22.00      22.00         250    96.9
## 2007-01-29     22.80     23.20    22.00      23.20         986    96.9
## 2007-01-31     23.60     24.00    23.60      24.00         125    96.9
## 2007-02-01     24.00     24.00    24.00      24.00         270    91.3
## 2007-02-02     23.60     24.00    23.60      24.00         729    91.3
## 2007-02-05     24.00     25.60    24.00      25.60         375    91.3
## 2007-02-06     24.40     24.40    24.40      24.40         142    91.3
## 2007-02-09     24.00     24.00    23.20      23.60        2690    91.3
## 2017-04-01      0.68      0.68     0.68       0.68        1000    97.0
```

```
## 2017-05-01      0.68      0.68     0.68       0.68        1000    97.1
## 2017-06-01      0.68      0.68     0.68       0.68        1000    95.1
## 2017-07-01      0.68      0.68     0.68       0.68        1000    93.4
## 2017-08-01      0.68      0.68     0.68       0.68        1000    96.8
## 2017-09-01      0.68      0.68     0.68       0.68        1000    95.1
```

d. then create a new series, named UMCSENT04, and replace the NAs using linear interpolation.

```
# we deem that it would be inappropriate to interpolate over
# more than a month's time.
UMCSENT04 <- xts(UMCSENT02)
UMCSENT04 <- round(na.approx(UMCSENT04, maxgap = 30), 2)
```

e. Print out some observations to ensure that your merge as well as the missing value imputation are done correctly.

```
rbind(head(UMCSENT04), UMCSENT04["2007-01-26/2007-02-09"], tail(UMCSENT04))
```

```
##            AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume UMCSENT
## 1978-01-01        NA        NA       NA         NA           NA   83.70
## 1978-02-01        NA        NA       NA         NA           NA   84.30
## 1978-03-01        NA        NA       NA         NA           NA   78.80
## 1978-04-01        NA        NA       NA         NA           NA   81.60
## 1978-05-01        NA        NA       NA         NA           NA   82.90
## 1978-06-01        NA        NA       NA         NA           NA   80.00
## 2007-01-26      22.0      22.0     22.0       22.0          250   92.38
## 2007-01-29      22.8      23.2     22.0       23.2          986   91.84
## 2007-01-31      23.6      24.0     23.6       24.0          125   91.48
## 2007-02-01      24.0      24.0     24.0       24.0          270   91.30
## 2007-02-02      23.6      24.0     23.6       24.0          729   91.20
## 2007-02-05      24.0      25.6     24.0       25.6          375   90.89
## 2007-02-06      24.4      24.4     24.4       24.4          142   90.78
## 2007-02-09      24.0      24.0     23.2       23.6         2690   90.47
## 2017-04-01        NA        NA       NA         NA           NA   97.00
## 2017-05-01        NA        NA       NA         NA           NA   97.10
## 2017-06-01        NA        NA       NA         NA           NA   95.10
## 2017-07-01        NA        NA       NA         NA           NA   93.40
## 2017-08-01        NA        NA       NA         NA           NA   96.80
## 2017-09-01        NA        NA       NA         NA           NA   95.10
```

We printed 20 ~ 30 samples for each set (5~6 at the beginning / the end, and the rest from the middle of the dataset that have values for AMAZ but lack values for UMCSENT apart from one in the middle (this allowed us to observe more easily the effects of imputing values).

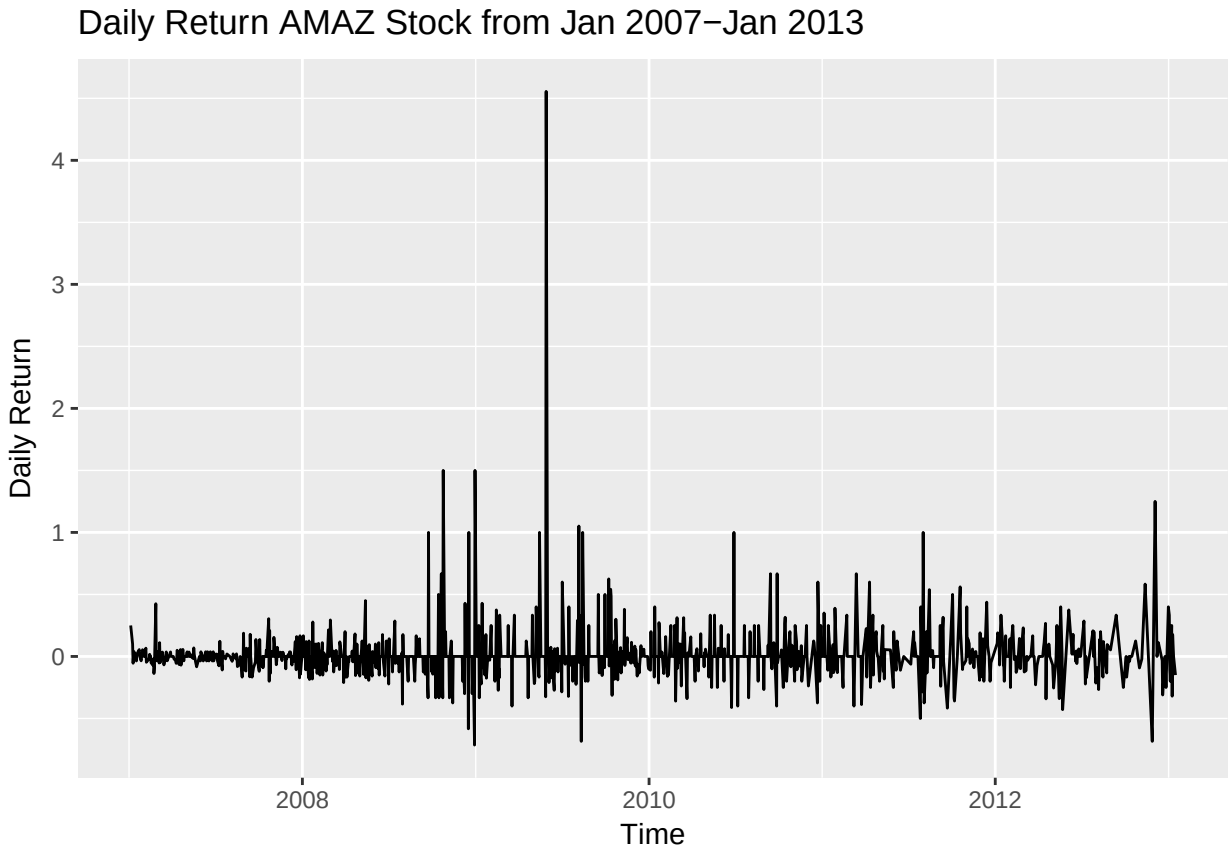
4. Calculate the daily return of the Amazon closing price (AMAZ.close), where daily return is defined as $(x(t) - x(t-1))/x(t-1)$. Plot the daily return series.

```
XT = AMAZ[, 4]
dXT = diff(XT)
daily.Return.AMAZ = dXT/lag(XT, k = 1)
df = cbind(XT, dXT, daily.Return.AMAZ)
colnames(df) <- c("AMAZ.close", "Delta", "Daily.Return")
rbind(head(df), tail(df))
```

```
##            AMAZ.close Delta Daily.Return
## 2007-01-03      16.00    NA           NA
## 2007-01-04      20.00  4.00   0.25000000
## 2007-01-08      22.00  2.00   0.10000000
```

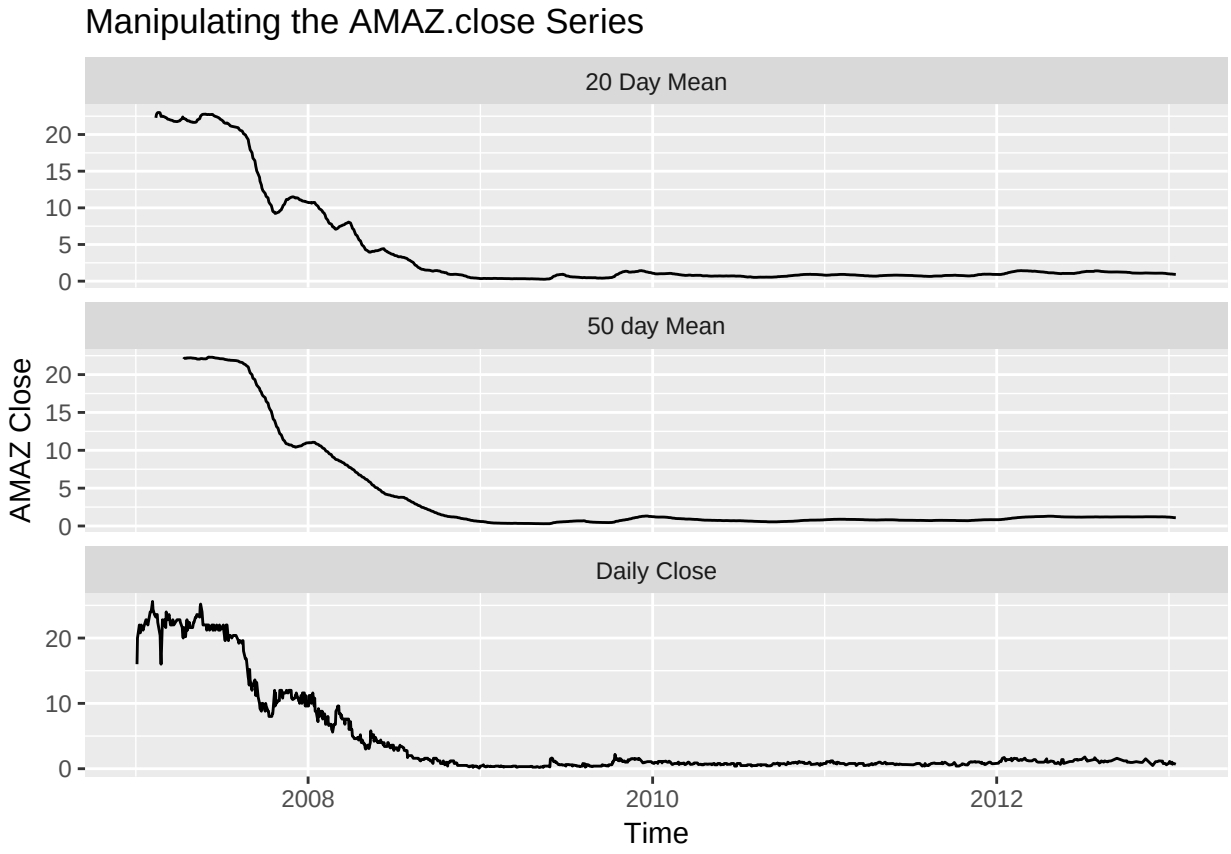
```
## 2007-01-09      20.80 -1.20  -0.05454545
## 2007-01-10      20.80  0.00   0.00000000
## 2007-01-11      21.60  0.80   0.03846154
## 2013-01-04       0.80 -0.20  -0.20000000
## 2013-01-07       1.00  0.20   0.25000000
## 2013-01-08       0.68 -0.32  -0.32000000
## 2013-01-09       0.80  0.12   0.17647059
## 2013-01-11       0.80  0.00   0.00000000
## 2013-01-15       0.68 -0.12  -0.15000000
```

```
library(ggfortify)
autoplot(daily.Return.AMAZ, main = "Daily Return AMAZ Stock from Jan 2007-Jan 2013") +
    ylab("Daily Return") + xlab("Time")
```



5. Create a 20-day and a 50-day rolling mean series from the AMAZ.close series.

```
AMAZ.close = AMAZ[, 4]
AMAZ.close.20.day.mean = rollapply(AMAZ.close, 20, FUN = mean,
    na.rm = TRUE)
AMAZ.close.50.day.mean = rollapply(AMAZ.close, 50, FUN = mean,
    na.rm = TRUE)
AMAZ.close.combined = cbind(AMAZ.close, AMAZ.close.20.day.mean,
    AMAZ.close.50.day.mean)
colnames(AMAZ.close.combined) = c("Daily Close", "20 Day Mean",
    "50 day Mean")
autoplot(AMAZ.close.combined, facets = T, main = "Manipulating the AMAZ.close Series") +
    ylab("AMAZ Close") + xlab("Time")
```


Manipulating the AMAZ.close Series
20 Day Mean
50 day Mean
Daily Close
AMAZ Close
0
5
10
15
20
2008
2010
2012
Time