W271 Section 3 Lab 3

Kiersten Henderson, Zhaoning Yu, Daghan Altas

11/11/2017

Task 5:

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

```
library(xts)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
wd <-"/Users/daghanaltas/Hacking/Berkeley/W271/Labs/Lab3_2017Fall"
setwd(wd)
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
```

```r
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 ...
```

```r
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

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

```
## Warning in strptime(x, format, tz = tz): unknown timezone 'default/America/
## Los_Angeles'
```

```r
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
```

```r
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)
rbind(head(AMAZ_UMCSENT_outer),AMAZ_UMCSENT_outer[sample(1:nrow(AMAZ_UMCSENT_outer),size = 20),], tail(
```

```
##            AMAZ.Open AMAZ.High AMAZ.Low AMAZ.Close AMAZ.Volume UMCSENT
## 1978-01-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    83.7
## 1978-02-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    84.3
## 1978-03-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    78.8
## 1978-04-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    81.6
## 1978-05-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    82.9
## 1978-06-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    80.0
## 1980-10-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    75.0
## 1989-11-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    90.9
## 1991-06-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    82.1
## 1996-03-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    93.7
## 2008-07-22        NA        NA       NA       3.20           0 -9999.0
## 2008-08-18      1.60      1.60     1.60       1.60          74 -9999.0
## 2008-09-17      1.60      1.60     1.20       1.20         498 -9999.0
## 2009-02-09        NA        NA       NA       0.40           0 -9999.0
## 2009-04-09        NA        NA       NA       0.32           0 -9999.0
## 2009-08-06      0.39      1.20     0.32       0.80       11688 -9999.0
## 2010-01-15        NA        NA       NA       1.12           0 -9999.0
## 2010-02-08      1.00      1.00     1.00       1.00         250 -9999.0
## 2011-03-07      0.80      0.80     0.80       0.80         500 -9999.0
## 2011-09-07      0.96      1.00     0.96       1.00        1250 -9999.0
## 2011-11-18      1.00      1.00     1.00       1.00         250 -9999.0
## 2012-03-20      1.36      1.40     1.36       1.40       12500 -9999.0
## 2012-04-10      1.00      1.04     1.00       1.00        4500 -9999.0
## 2012-08-17      1.00      1.20     1.00       1.20         188 -9999.0
## 2013-01-02      1.12      1.12     1.12       1.12           2 -9999.0
## 2016-03-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    91.0
## 2017-04-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    97.0
## 2017-05-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    97.1
## 2017-06-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    95.1
## 2017-07-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    93.4
## 2017-08-01  -9999.00  -9999.00 -9999.00   -9999.00       -9999    96.8
## 2017-09-01  -9999.00  -9999.00 -9999.00   -9999.00       -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[sample(1:nrow(UMCSENT02),size = 20),], 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
## 1983-02-01        NA        NA       NA         NA          NA    74.6
## 1989-09-01        NA        NA       NA         NA          NA    95.8
```

```
## 1991-03-01       NA      NA      NA      NA       NA   87.7
## 1994-04-01       NA      NA      NA      NA       NA   92.6
## 2002-07-01       NA      NA      NA      NA       NA   88.1
## 2002-09-01       NA      NA      NA      NA       NA   86.1
## 2007-02-23     18.80   18.80   16.00   16.00      375     NA
## 2007-07-19     20.40   20.40   20.40   20.40       69     NA
## 2008-05-02      3.00    3.00    3.00    3.00       25     NA
## 2008-05-19      5.20    5.20    5.20    5.20      250     NA
## 2008-08-01       NA      NA      NA    2.00        0   63.0
## 2010-01-06      0.96    0.96    0.96    0.96     3725     NA
## 2010-03-19      1.00    1.00    1.00    1.00      100     NA
## 2010-06-01       NA      NA      NA    0.64        0   76.0
## 2011-02-02      0.80    0.80    0.80    0.80      125     NA
## 2011-05-25      0.88    0.88    0.80    0.80     1750     NA
## 2011-10-20      0.88    0.88    0.76    0.76      305     NA
## 2013-01-01       NA      NA      NA      NA       NA   73.8
## 2013-10-01       NA      NA      NA      NA       NA   73.2
## 2015-02-01       NA      NA      NA      NA       NA   95.4
## 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)
UMCSENT03 <- na.locf(UMCSENT03, na.rm = TRUE, fromLast = FALSE)
rbind(head(UMCSENT03),UMCSENT03[sample(1:nrow(UMCSENT03),size = 20),], tail(UMCSENT03))
```

```
##            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
## 2007-10-04      9.60      9.60     9.60       9.60         250    80.9
## 2007-11-01     10.40     10.40    10.40      10.40           0    76.1
## 2008-01-03     11.20     11.20    11.20      11.20         125    78.4
## 2008-01-18     10.80     10.80     8.80       8.80           0    78.4
## 2008-02-12      8.00      8.00     8.00       8.00          16    70.8
## 2008-08-14      1.60      1.60     1.60       1.60           0    63.0
## 2009-02-09      0.40      0.40     0.40       0.40           0    56.3
## 2009-02-23      0.32      0.32     0.32       0.32         528    56.3
## 2009-03-31      0.32      0.32     0.32       0.32           0    57.3
## 2009-04-22      0.24      0.24     0.24       0.24           0    65.1
## 2010-04-19      0.72      0.72     0.60       0.60           0    72.2
## 2010-05-25      0.60      0.60     0.60       0.60         125    73.6
## 2010-06-21      0.80      0.80     0.80       0.80           0    76.0
## 2010-07-09      0.48      0.48     0.48       0.48           0    67.8
## 2011-03-01      0.80      0.80     0.80       0.80         750    67.5
## 2011-04-04      0.50      0.60     0.50       0.60        5000    69.8
## 2011-04-28      1.00      1.00     1.00       1.00         195    69.8
```

```
## 2011-09-29        0.55      0.64      0.55      0.64       395    59.5
## 2011-12-19        1.00      1.00      1.00      1.00         2    69.9
## 2012-01-25        1.40      1.40      1.40      1.40       125    75.0
## 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.

```r
UMCSENT04 <- xts(UMCSENT02)
UMCSENT04 <- na.approx(UMCSENT04, maxgap= 30)
rbind(head(UMCSENT04),UMCSENT04[sample(1:nrow(UMCSENT04),size = 20),], tail(UMCSENT04))
```

```
##              AMAZ.Open  AMAZ.High   AMAZ.Low AMAZ.Close AMAZ.Volume
## 1978-01-01         NA         NA         NA         NA          NA
## 1978-02-01         NA         NA         NA         NA          NA
## 1978-03-01         NA         NA         NA         NA          NA
## 1978-04-01         NA         NA         NA         NA          NA
## 1978-05-01         NA         NA         NA         NA          NA
## 1978-06-01         NA         NA         NA         NA          NA
## 1987-11-01         NA         NA         NA         NA          NA
## 1999-11-01         NA         NA         NA         NA          NA
## 2002-03-01         NA         NA         NA         NA          NA
## 2007-04-01      22.96 22.9600000 22.8000000      22.80       154.6
## 2007-09-21       9.80 10.0000000  9.2000000       9.20       936.0
## 2008-03-20       5.60  7.6000000  5.6000000       7.60       526.0
## 2008-04-21       5.20  5.2000000  5.2000000       5.20      1500.0
## 2008-07-01       3.60  3.6000000  2.8000000       2.80       150.0
## 2008-11-18       0.50  0.5000000  0.5000000       0.50         0.0
## 2008-12-24       0.28  0.2800000  0.2800000       0.28       220.0
## 2009-01-16       0.40  0.4000000  0.3800000       0.40         0.0
## 2009-06-04       1.52  1.5200000  0.7200000       1.20      8390.0
## 2009-08-06       0.39  1.2000000  0.3200000       0.80     11688.0
## 2010-04-15       0.72  0.7200000  0.6000000       0.60      2725.0
## 2010-04-27       0.66  0.6600000  0.6600000       0.66       125.0
## 2010-05-06       0.66  0.6600000  0.6600000       0.72         0.0
## 2010-07-13       0.48  0.5314286  0.4457143       0.48         0.0
## 2011-04-18       0.80  0.8000000  0.8000000       0.80       112.0
## 2011-11-03       0.84  0.8400000  0.8400000       0.84      1125.0
## 2012-04-02       0.79  1.0000000  0.6800000       1.00       750.0
## 2017-04-01         NA         NA         NA         NA          NA
## 2017-05-01         NA         NA         NA         NA          NA
## 2017-06-01         NA         NA         NA         NA          NA
## 2017-07-01         NA         NA         NA         NA          NA
## 2017-08-01         NA         NA         NA         NA          NA
## 2017-09-01         NA         NA         NA         NA          NA
##              UMCSENT
## 1978-01-01  83.70000
## 1978-02-01  84.30000
## 1978-03-01  78.80000
## 1978-04-01  81.60000
## 1978-05-01  82.90000
```

```
## 1978-06-01  80.00000
## 1987-11-01  83.10000
## 1999-11-01 107.20000
## 2002-03-01  95.70000
## 2007-04-01  87.10000
## 2007-09-21  81.73333
## 2008-03-20  65.27097
## 2008-04-21  60.73333
## 2008-07-01  61.20000
## 2008-11-18  58.02000
## 2008-12-24  60.91613
## 2009-01-16  58.82903
## 2009-06-04  70.32000
## 2009-08-06  66.95806
## 2010-04-15  72.85333
## 2010-04-27  73.41333
## 2010-05-06  73.98710
## 2010-07-13  68.22581
## 2011-04-18  72.35000
## 2011-11-03  64.11333
## 2012-04-02  76.49667
## 2017-04-01  97.00000
## 2017-05-01  97.10000
## 2017-06-01  95.10000
## 2017-07-01  93.40000
## 2017-08-01  96.80000
## 2017-09-01  95.10000
```

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

I have printed 30 samples for each set (5 at the beginning, 20 random samples and 5 at the end).

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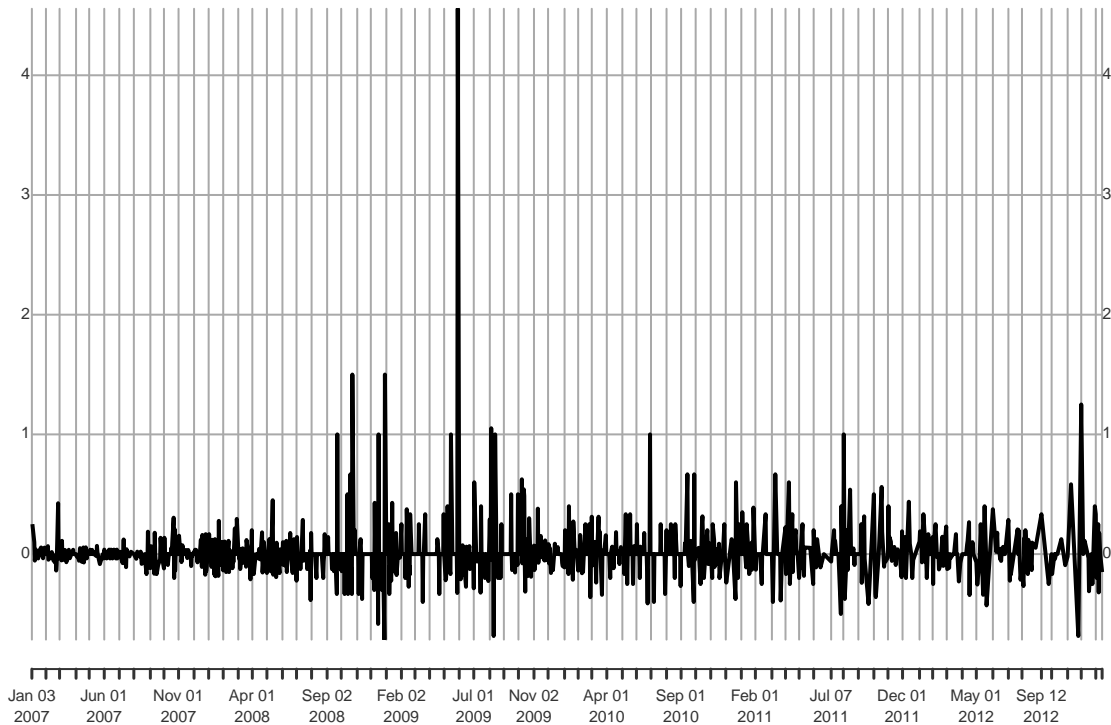
```
XT = AMAZ[,4]
dXT = diff(XT)
daily.Return.AMAZ = dXT/lag(XT, k = 1)
df = cbind(XT, dXT, daily.Return.AMAZ)
colnames(df) <- c("AMAZ.close", "Delta", "Daily.Return")
rbind(head(df),tail(df))
```

```
##            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
```

```r
plot(daily.Return.AMAZ)
```

daily.Return.AMAZ   2007−01−03 / 2013−01−15

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

```r
library(ggfortify)
```

```
## Loading required package: ggplot2
```

```r
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 = F)
```

```
## Warning: Removed 68 rows containing missing values (geom_path).
```