

W271 Section 3 Lab 3

Kiersten Henderson, Zhaoning Yu, Daghan Altas

11/11/2017

Task 5:

1. Read AMAZ.csv and UMCSSENT.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")
UMCSSENT_df <- read.csv("./UMCSSENT.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

```
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)
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
## 1988-09-01	-9999.00	-9999.00	-9999.00	-9999.00	-9999	97.3
## 1989-09-01	-9999.00	-9999.00	-9999.00	-9999.00	-9999	95.8
## 1999-05-01	-9999.00	-9999.00	-9999.00	-9999.00	-9999	106.8
## 2007-11-21	12.00	12.00	12.00	12.00	25	-9999.0
## 2008-03-31	6.00	7.20	6.00	7.20	455	-9999.0
## 2008-05-20	4.40	4.40	4.20	4.20	375	-9999.0
## 2008-08-14	NA	NA	NA	1.60	0	-9999.0
## 2009-01-30	0.40	0.40	0.32	0.32	8250	-9999.0
## 2009-02-01	-9999.00	-9999.00	-9999.00	-9999.00	-9999	56.3
## 2009-06-08	1.16	1.16	1.16	1.16	525	-9999.0
## 2009-06-19	0.60	0.60	0.60	0.60	150	-9999.0
## 2009-09-02	NA	NA	NA	0.40	0	-9999.0
## 2010-03-05	0.80	0.80	0.80	0.80	125	-9999.0
## 2010-05-01	-9999.00	-9999.00	-9999.00	-9999.00	-9999	73.6
## 2010-05-04	0.72	0.72	0.72	0.72	16	-9999.0
## 2010-09-28	0.80	0.80	0.80	0.80	50	-9999.0
## 2011-01-18	1.20	1.20	1.20	1.20	625	-9999.0
## 2012-01-27	1.20	1.40	1.20	1.40	625	-9999.0
## 2012-02-23	1.28	1.28	1.28	1.28	150	-9999.0
## 2013-05-01	-9999.00	-9999.00	-9999.00	-9999.00	-9999	84.5
## 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
## 1985-08-01	NA	NA	NA	NA	NA	92.4
## 1992-06-01	NA	NA	NA	NA	NA	80.4
## 1994-02-01	NA	NA	NA	NA	NA	93.2
## 1998-01-01	NA	NA	NA	NA	NA	106.6

## 2006-09-01	NA	NA	NA	NA	NA	85.4
## 2007-07-02	21.20	21.60	21.20	21.60	17	NA
## 2007-08-21	16.80	16.80	16.80	16.80	125	NA
## 2007-09-01	NA	NA	NA	NA	NA	83.4
## 2008-04-28	4.00	4.00	3.80	3.80	775	NA
## 2009-06-22	0.64	0.64	0.64	0.64	125	NA
## 2009-06-24	NA	NA	NA	0.56	0	NA
## 2009-07-09	0.60	0.60	0.60	0.60	100	NA
## 2009-10-14	2.00	2.20	1.60	2.04	10058	NA
## 2009-11-20	1.56	1.56	1.36	1.56	7110	NA
## 2009-11-25	1.40	1.56	1.40	1.56	2625	NA
## 2009-12-24	NA	NA	NA	1.08	0	NA
## 2009-12-29	0.88	1.08	0.80	1.08	2638	NA
## 2010-03-22	1.00	1.00	0.64	0.66	175	NA
## 2010-04-08	0.80	0.80	0.64	0.64	875	NA
## 2010-08-12	NA	NA	NA	0.60	0	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

```
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
## 2007-06-21	22.00	22.00	22.00	22.00	250	85.3
## 2008-04-11	4.80	4.80	4.60	4.60	0	62.6
## 2008-05-12	3.60	4.40	3.60	4.00	1650	59.8
## 2008-05-28	4.00	4.20	4.00	4.20	112	59.8
## 2008-09-01	1.20	1.40	1.20	1.40	0	70.3
## 2008-10-29	0.60	1.20	0.60	1.20	1005	57.6
## 2009-02-09	0.40	0.40	0.40	0.40	0	56.3
## 2009-03-04	0.32	0.32	0.32	0.32	0	57.3
## 2009-03-17	0.40	0.40	0.40	0.40	0	57.3
## 2009-04-23	0.24	0.24	0.24	0.24	0	65.1
## 2009-08-19	0.60	0.60	0.50	0.50	625	65.7
## 2010-07-27	0.48	0.60	0.40	0.60	0	67.8
## 2010-10-21	0.72	0.80	0.72	0.80	492	67.7
## 2011-01-28	0.72	1.00	0.72	1.00	1625	74.2
## 2011-09-13	1.00	1.00	1.00	1.00	140	59.5
## 2011-11-02	0.84	0.84	0.84	0.84	1000	63.7
## 2012-01-30	1.40	1.40	1.40	1.40	250	75.0
## 2012-02-23	1.28	1.28	1.28	1.28	150	75.3
## 2012-03-21	1.40	1.40	1.40	1.40	1250	76.2

```
## 2016-07-01      0.68      0.68      0.68      0.68      1000      90.0
## 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.

```
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
## 1987-06-01         NA         NA         NA         NA         NA
## 1999-05-01         NA         NA         NA         NA         NA
## 2002-11-01         NA         NA         NA         NA         NA
## 2006-01-01         NA         NA         NA         NA         NA
## 2007-01-09 22.0000000 22.0000000 20.8000000      20.80         356
## 2007-07-05 22.0000000 22.0000000 22.0000000      22.00           8
## 2007-11-21 12.0000000 12.0000000 12.0000000      12.00          25
## 2008-01-24  8.8000000  8.8000000  8.8000000       8.80          18
## 2008-02-11  7.2000000  7.2000000  7.2000000       7.20         169
## 2008-03-19  6.0000000  6.8000000  6.0000000       6.80         505
## 2008-04-08  5.2000000  5.2000000  4.6000000       4.60         430
## 2008-06-18  4.0000000  4.0000000  4.0000000       4.00         190
## 2009-07-06  0.6133333  0.6133333  0.6133333       0.64           0
## 2009-11-11  1.2000000  1.4000000  1.2000000       1.28        5803
## 2010-04-05  0.8800000  0.8800000  0.8800000       0.88         762
## 2010-06-16  0.7600000  0.7600000  0.7600000       0.68           0
## 2010-06-24  0.4000000  0.4000000  0.4000000       0.40         833
## 2010-06-28  0.7200000  0.7200000  0.7200000       0.40           0
## 2011-01-14  1.0000000  1.2000000  0.9600000       1.20        1766
## 2012-08-10  1.2000000  1.2000000  1.2000000       1.20         125
## 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
## 1987-06-01 91.50000
```

```
## 1999-05-01 106.80000
## 2002-11-01 84.20000
## 2006-01-01 91.20000
## 2007-01-09 95.45484
## 2007-07-05 89.49677
## 2007-11-21 75.69972
## 2008-01-24 72.76129
## 2008-02-11 70.35172
## 2008-03-19 65.49744
## 2008-04-08 61.94667
## 2008-06-18 59.12000
## 2009-07-06 65.95161
## 2009-11-11 69.10472
## 2010-04-05 72.38667
## 2010-06-16 71.90000
## 2010-06-24 69.71333
## 2010-06-28 68.62000
## 2011-01-14 75.58387
## 2012-08-10 75.46129
## 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
```

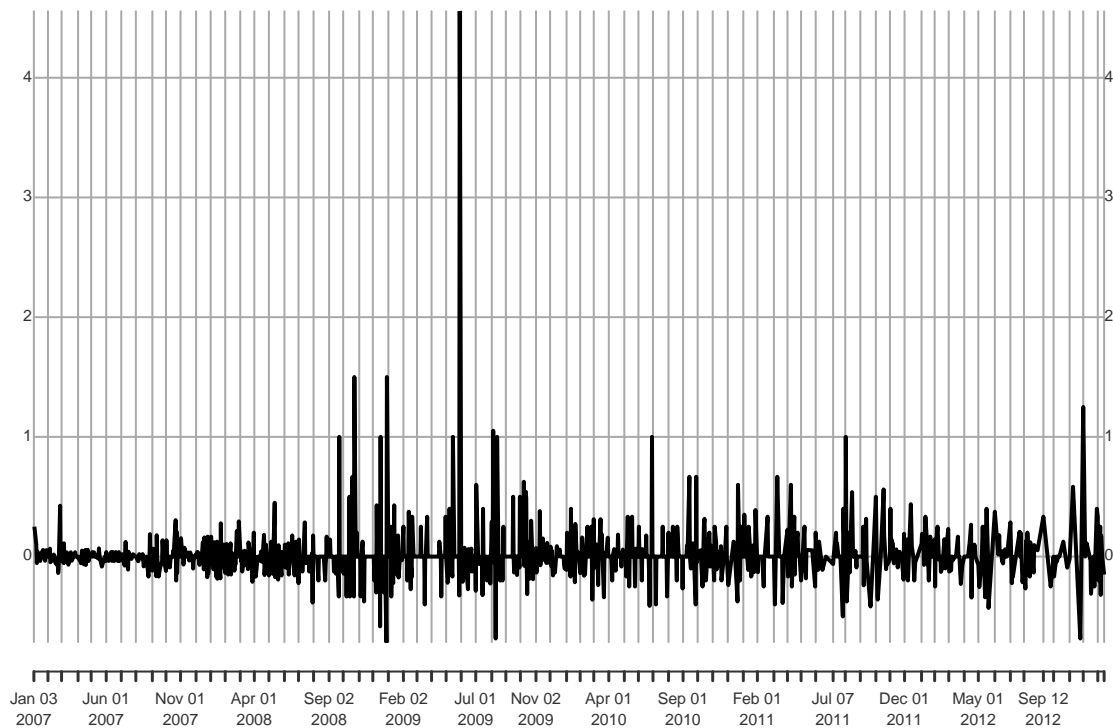
e. Print out some observations to ensure that your merge as well as the missing value imputation are done correctly. I have printed 30 samples for each set (5 at the beginning, 20 random samples and 5 at the end).

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
```

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



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

```
library(ggfortify)
```

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

```
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).
```