

Project 12 -- Seyi Ogunmodede

Instructor Help: Dr. Mark

- Help with figuring out how to write the programs function using the videos.

Collaboration:

-

Question 1

```
In [1]: options(jupyter.rich_display = F)
```

```
In [2]: options(repr.matrix.max.cols=100,  
             repr.matrix.max.rows=200)
```

```
In [3]: library(data.table)
```

```
In [4]: state <-fread("/anvil/projects/tdm/data/zillow/State_time_series.csv")  
# fread is used to read in the data.
```

```
In [5]: dim(state)
```

```
[1] 13212     82
```

```
In [6]: library(lubridate)
```

```
Warning message in system("timedatectl", intern = TRUE):  
"running command 'timedatectl' had status 1"
```

```
Attaching package: 'lubridate'
```

```
The following objects are masked from 'package:data.table':
```

```
hour, isoweek, mday, minute, month, quarter, second, wday, week,  
yday, year
```

```
The following objects are masked from 'package:base':
```

```
date, intersect, setdiff, union
```

```
In [7]: head(state)
```

	Date	RegionName	DaysOnZillow_AllHomes
1	1996-04-30	Alabama	NA
2	1996-04-30	Arizona	NA
3	1996-04-30	Arkansas	NA
4	1996-04-30	California	NA
5	1996-04-30	Colorado	NA
6	1996-04-30	Connecticut	NA
	InventorySeasonallyAdjusted_AllHomes	InventoryRaw_AllHomes	
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
	MedianListingPricePerSqft_1Bedroom	MedianListingPricePerSqft_2Bedroom	
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
	MedianListingPricePerSqft_3Bedroom	MedianListingPricePerSqft_4Bedroom	
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
	MedianListingPricePerSqft_5BedroomOrMore	MedianListingPricePerSqft_AllHomes	
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
	MedianListingPricePerSqft_CondoCoop	MedianListingPricePerSqft_DuplexTriplex	
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
	MedianListingPricePerSqft_SingleFamilyResidence	MedianListingPrice_1Bedroom	
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
	MedianListingPrice_2Bedroom	MedianListingPrice_3Bedroom	
1	NA	NA	
2	NA	NA	
3	NA	NA	
4	NA	NA	
5	NA	NA	
6	NA	NA	
	MedianListingPrice_4Bedroom	MedianListingPrice_5BedroomOrMore	
1	NA	NA	
2	NA	NA	
3	NA	NA	

```
4 NA NA
5 NA NA
6 NA NA
    MedianListingPrice_AllHomes MedianListingPrice_CondoCoop
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianListingPrice_DuplexTriplex MedianListingPrice_SingleFamilyResidence
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianPctOfPriceReduction_AllHomes MedianPctOfPriceReduction_CondoCoop
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianPctOfPriceReduction_SingleFamilyResidence MedianPriceCutDollar_AllHomes
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianPriceCutDollar_CondoCoop MedianPriceCutDollar_SingleFamilyResidence
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPricePerSqft_1Bedroom MedianRentalPricePerSqft_2Bedroom
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPricePerSqft_3Bedroom MedianRentalPricePerSqft_4Bedroom
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPricePerSqft_5BedroomOrMore MedianRentalPricePerSqft_AllHomes
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPricePerSqft_CondoCoop MedianRentalPricePerSqft_DuplexTriplex
```

```
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPricePerSqft_MultiFamilyResidence5PlusUnits
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
    MedianRentalPricePerSqft_SingleFamilyResidence
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
    MedianRentalPricePerSqft_Studio MedianRentalPrice_1Bedroom
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPrice_2Bedroom MedianRentalPrice_3Bedroom
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPrice_4Bedroom MedianRentalPrice_5BedroomOrMore
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPrice_AllHomes MedianRentalPrice_CondoCoop
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
    MedianRentalPrice_DuplexTriplex
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
    MedianRentalPrice_MultiFamilyResidence5PlusUnits
1 NA
2 NA
3 NA
4 NA
```

```
5 NA
6 NA
  MedianRentalPrice_SingleFamilyResidence MedianRentalPrice_Studio
1 NA                               NA
2 NA                               NA
3 NA                               NA
4 NA                               NA
5 NA                               NA
6 NA                               NA
  ZHVIPerSqft_AllHomes PctOfHomesDecreasingInValues_AllHomes
1 50                                NA
2 62                                NA
3 42                                NA
4 102                               NA
5 82                                NA
6 85                                NA
  PctOfHomesIncreasingInValues_AllHomes PctOfHomesSellingForGain_AllHomes
1 NA                                NA
2 NA                                NA
3 NA                                NA
4 NA                                NA
5 NA                                NA
6 NA                                NA
  PctOfHomesSellingForLoss_AllHomes
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
  PctOfListingsWithPriceReductionsSeasAdj_AllHomes
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
  PctOfListingsWithPriceReductionsSeasAdj_CondoCoop
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
  PctOfListingsWithPriceReductionsSeasAdj_SingleFamilyResidence
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
  PctOfListingsWithPriceReductions_AllHomes
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
  PctOfListingsWithPriceReductions_CondoCoop
1 NA
```

```

2 NA
3 NA
4 NA
5 NA
6 NA
  PctOfListingsWithPriceReductions_SingleFamilyResidence
1 NA
2 NA
3 NA
4 NA
5 NA
6 NA
  PriceToRentRatio_AllHomes Sale_Counts Sale_Counts_Seas_Adj Sale_Prices
1 NA NA NA NA
2 NA NA NA NA
3 NA NA NA NA
4 NA NA NA NA
5 NA NA NA NA
6 NA NA NA NA
  ZHVI_1bedroom ZHVI_2bedroom ZHVI_3bedroom ZHVI_4bedroom ZHVI_5BedroomOrMore
1 61500 48900 78200 146500 206300
2 59200 86400 96100 128400 190500
3 53000 54500 76800 135100 186000
4 93700 123400 150900 196100 265300
5 77800 97500 129000 176100 212900
6 64700 97000 130400 194800 299800
  ZHVI_AllHomes ZHVI_BottomTier ZHVI_CondoCoop ZHVI_MiddleTier
1 79500 45600 99500 79500
2 103600 67100 78900 103600
3 64400 38400 70300 64400
4 157900 95100 136100 157900
5 128100 82700 99400 128100
6 132000 83700 85000 132000
  ZHVI_SingleFamilyResidence ZHVI_TopTier ZRI_AllHomes
1 79000 140200 NA
2 107500 168700 NA
3 64500 115200 NA
4 162000 270600 NA
5 133600 209300 NA
6 141000 231600 NA
  ZRI_AllHomesPlusMultifamily ZriPerSqft_AllHomes
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA
  Zri_MultiFamilyResidenceRental Zri_SingleFamilyResidenceRental
1 NA NA
2 NA NA
3 NA NA
4 NA NA
5 NA NA
6 NA NA

```

In [8]: `class(state$date)`

```
[1] "IDate" "Date"
```

In [9]: `class(Date)`

```
[1] "function"
```

```
In [10]: typeof(state$date)
```

```
[1] "integer"
```

```
In [11]: str(state)
```

```
# this helps to get all the types
```

```
Classes 'data.table' and 'data.frame': 13212 obs. of  82 variables:
 $ Date                               : IDate, format: "199
6-04-30" "1996-04-30" ...
 $ RegionName                         : chr  "Alabama" "Ari
zona" "Arkansas" "California" ...
 $ DaysOnZillow_AllHomes              : num  NA NA NA NA NA
NA NA NA NA ...
 $ InventorySeasonallyAdjusted_AllHomes: int  NA NA NA NA NA
NA NA NA NA ...
 $ InventoryRaw_AllHomes              : int  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_1Bedroom: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_2Bedroom: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_3Bedroom: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_4Bedroom: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_5BedroomOrMore: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_AllHomes: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_CondoCoop: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_DuplexTriplex: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPricePerSqft_SingleFamilyResidence: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_1Bedroom         : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_2Bedroom         : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_3Bedroom         : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_4Bedroom         : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_5BedroomOrMore   : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_AllHomes        : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_CondoCoop       : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_DuplexTriplex  : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianListingPrice_SingleFamilyResidence: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianPctOfPriceReduction_AllHomes: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianPctOfPriceReduction_CondoCoop: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianPctOfPriceReduction_SingleFamilyResidence: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianPriceCutDollar_AllHomes      : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianPriceCutDollar_CondoCoop     : num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianPriceCutDollar_SingleFamilyResidence: num  NA NA NA NA NA
NA NA NA NA ...
 $ MedianRentalPricePerSqft_1Bedroom  : num  NA NA NA NA NA
```

NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_2Bedroom	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_3Bedroom	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_4Bedroom	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_5BedroomOrMore	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_AllHomes	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_CondoCoop	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_DuplexTriplex	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_MultiFamilyResidence5PlusUnits	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_SingleFamilyResidence	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPricePerSqft_Studio	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_1Bedroom	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_2Bedroom	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_3Bedroom	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_4Bedroom	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_5BedroomOrMore	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_AllHomes	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_CondoCoop	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_DuplexTriplex	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_MultiFamilyResidence5PlusUnits	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_SingleFamilyResidence	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ MedianRentalPrice_Studio	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ ZHVIPerSqft_AllHomes	: int 50 62 42 102 8
2 85 71 56 55 185 ...	
\$ PctOfHomesDecreasingInValues_AllHomes	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ PctOfHomesIncreasingInValues_AllHomes	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ PctOfHomesSellingForGain_AllHomes	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ PctOfHomesSellingForLoss_AllHomes	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ PctOfListingsWithPriceReductionsSeasAdj_AllHomes	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ PctOfListingsWithPriceReductionsSeasAdj_CondoCoop	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ PctOfListingsWithPriceReductionsSeasAdj_SingleFamilyResidence	: num NA NA NA NA NA
NA NA NA NA NA ...	
\$ PctOfListingsWithPriceReductionsAllHomes	: num NA NA NA NA NA

```

NA NA NA NA NA ...
$ PctOfListingsWithPriceReductions_CondoCoop : num NA NA NA NA NA
NA NA NA NA NA ...
$ PctOfListingsWithPriceReductions_SingleFamilyResidence : num NA NA NA NA NA
NA NA NA NA NA ...
$ PriceToRentRatio_AllHomes : num NA NA NA NA NA
NA NA NA NA NA ...
$ Sale_Counts : num NA NA NA NA NA
NA NA NA NA NA ...
$ Sale_Counts_Seas_Adj : num NA NA NA NA NA
NA NA NA NA NA ...
$ Sale_Prices : num NA NA NA NA NA
NA NA NA NA NA ...
$ ZHVI_1bedroom : int 61500 59200 53
000 93700 77800 64700 90100 45400 74900 152300 ...
$ ZHVI_2bedroom : int 48900 86400 54
500 123400 97500 97000 88200 65400 64700 186600 ...
$ ZHVI_3bedroom : int 78200 96100 76
800 150900 129000 130400 103500 89100 88000 231800 ...
$ ZHVI_4bedroom : int 146500 128400
135100 196100 176100 194800 157800 133600 149700 303400 ...
$ ZHVI_5BedroomOrMore : int 206300 190500
186000 265300 212900 299800 176100 199900 212800 345500 ...
$ ZHVI_AllHomes : int 79500 103600 6
4400 157900 128100 132000 106800 86300 92000 227400 ...
$ ZHVI_BottomTier : int 45600 67100 38
400 95100 82700 83700 77200 52500 57200 144500 ...
$ ZHVI_CondoCoop : int 99500 78900 70
300 136100 99400 85000 NA 70600 89300 177000 ...
$ ZHVI_MiddleTier : int 79500 103600 6
4400 157900 128100 132000 106800 86300 92000 227400 ...
$ ZHVI_SingleFamilyResidence : int 79000 107500 6
4500 162000 133600 141000 107400 92100 92400 262600 ...
$ ZHVI_TopTier : int 140200 168700
115200 270600 209300 231600 161600 155300 163900 374700 ...
$ ZRI_AllHomes : int NA NA NA NA NA
NA NA NA NA ...
$ ZRI_AllHomesPlusMultifamily : int NA NA NA NA NA
NA NA NA NA ...
$ ZriPerSqft_AllHomes : num NA NA NA NA NA
NA NA NA NA ...
$ Zri_MultiFamilyResidenceRental : int NA NA NA NA NA
NA NA NA NA ...
$ Zri_SingleFamilyResidenceRental : int NA NA NA NA NA
NA NA NA NA ...
- attr(*, ".internal.selfref")=<externalptr>

```

In [12]: `head(state$date)`

```
[1] "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30"
[6] "1996-04-30"
```

Go ahead and read in the dataset as states

Find the class and the type of the column named Date

Are there multiple functions that will return the same or similar information? There are different output from each command use.

Question 2

```
In [13]: library(lubridate)
```

```
In [14]: head(state$date)
```

```
[1] "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30"
[6] "1996-04-30"
```

```
In [15]: class(head(state$date))
# before we have this.
```

```
[1] "IDate" "Date"
```

```
In [16]: ymd(head(state$date))
# running ymd function
```

```
[1] "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30"
[6] "1996-04-30"
```

```
In [17]: class(ymd(head(state$date)))
# but after introducing ymd it changes
# to know the class it belongs to
```

```
[1] "Date"
```

```
In [18]: state$newdates <- ymd(state$date)
# to make a whole new column call newdates
```

```
In [19]: head(state$newdates)
# i just got the date now and the date class
```

```
[1] "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30"
[6] "1996-04-30"
```

```
In [22]: wday(head(state$newdates),
label=TRUE)
```

```
[1] Tue Tue Tue Tue Tue
Levels: Sun < Mon < Tue < Wed < Thu < Fri < Sat
```

```
In [23]: tail(state$newdates)
```

```
[1] "2017-12-31" "2017-12-31" "2017-12-31" "2017-12-31" "2017-12-31"
[6] "2017-12-31"
```

```
In [24]: wday(tail(state$newdates),
label=TRUE)
```

```
[1] Sun Sun Sun Sun Sun
Levels: Sun < Mon < Tue < Wed < Thu < Fri < Sat
```

```
In [25]: state$day_of_the_week <- wday(state$newdates,
label=TRUE)
```

```
In [26]: table(state$day_of_the_week)
```

Sun	Mon	Tue	Wed	Thu	Fri	Sat
1873	1926	1871	1874	1828	1919	1921

```
In [43]: state$day_of_the_week_as_number <- wday(state$newdates)
```

```
In [28]: table(state$day_of_the_week_as_number )
# Sunday is day1, Monday is day2,.....Saturday is day7
```

	1	2	3	4	5	6	7
1873	1926	1871	1874	1828	1919	1921	

In Project 11, we had to convert the Date column to a month, day, year format. Now convert the column Date into values from the class Date. (You can use lubridate to do so.)

What do you think about the methods you have learned (so far) to convert dates?
but after introducing ymd it changes to know the class it belongs to

Create a new column in your data.frame states named day_of_the_week that shows (Sunday-Saturday).

Lets create another column in the data.frame states that shows the days of the week as numbers.

Question 3

```
In [29]: head (state$DaysOnZillow_AllHomes)
# it has Lot of NA's
```

```
[1] NA NA NA NA NA NA
```

```
In [30]: tail (state$DaysOnZillow_AllHomes)
# it has Lot of NA's
```

```
[1] NA NA NA NA NA NA
```

```
In [31]: summary(state$DaysOnZillow_AllHomes)
# then Look for the summary
# this shows different ranges
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	NA's
49.25	90.25	108.50	110.12	126.75	251.62	8367

```
In [44]: head(state$newdates)
# Look at the head of the new dates we created.
```

```
[1] "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30" "1996-04-30"
[6] "1996-04-30"
```

```
In [46]: month(head(state$newdates))
# to pull out the month that are there
```

```
[1] 4 4 4 4 4 4
```

```
In [34]: state$month <- month((state$newdates))
# to make a month column that has all the months entries
```

```
In [35]: tapply(state$DaysOnZillow_AllHomes,state$month, mean )
# do a tapply Looking at the number of day on Zillow all homes,
```

```
break it down into months and then find the average
# it has lot of NA's
```

```
1 2 3 4 5 6 7 8 9 10 11 12
NA NA
```

In [36]:

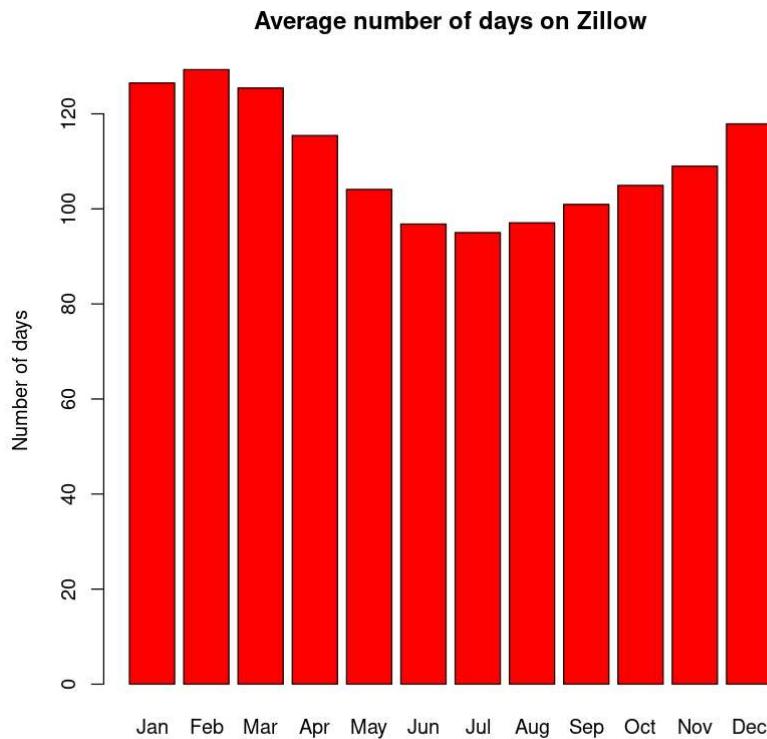
```
tapply(state$DaysOnZillow_AllHomes,state$month, mean,
       na.rm= T )
# to remove NA's
# this shows month by month, how many days on Zillow there are on average.
```

1	2	3	4	5	6	7	8
126.48836	129.29534	125.43444	115.42862	104.08793	96.81260	95.00980	97.07047
9	10	11	12				
100.94026	104.93689	108.99142	117.88294				

In [37]:

```
barplot(tapply(state$DaysOnZillow_AllHomes,state$month,
               mean, na.rm= T), names.arg= month.abb, col= "red",
               main= "Average number of days on Zillow",
               ylab= "Number of days")

# names there be the month of abbreviation (names.arg= month.abb)
```



We want to see if there is a better month(s) for putting our house on the market?

Use tapply to compare the average DaysOnZillow_AllHomes for all months. for the winter and fall months there is low sales while high sales in summer months.

Make a barplot showing our results.

Question 4

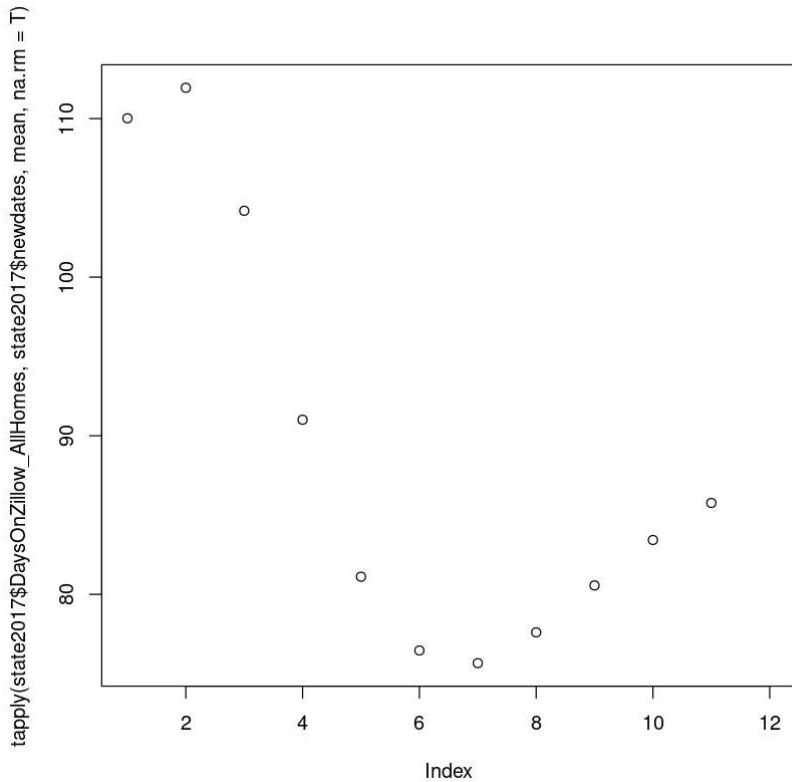
```
In [38]: state$year <- year((state$newdates))
# make a year column first
```

```
In [39]: state2017 <- subset(state, year==2017)
# take the state data frame and break
# it according to year there.
```

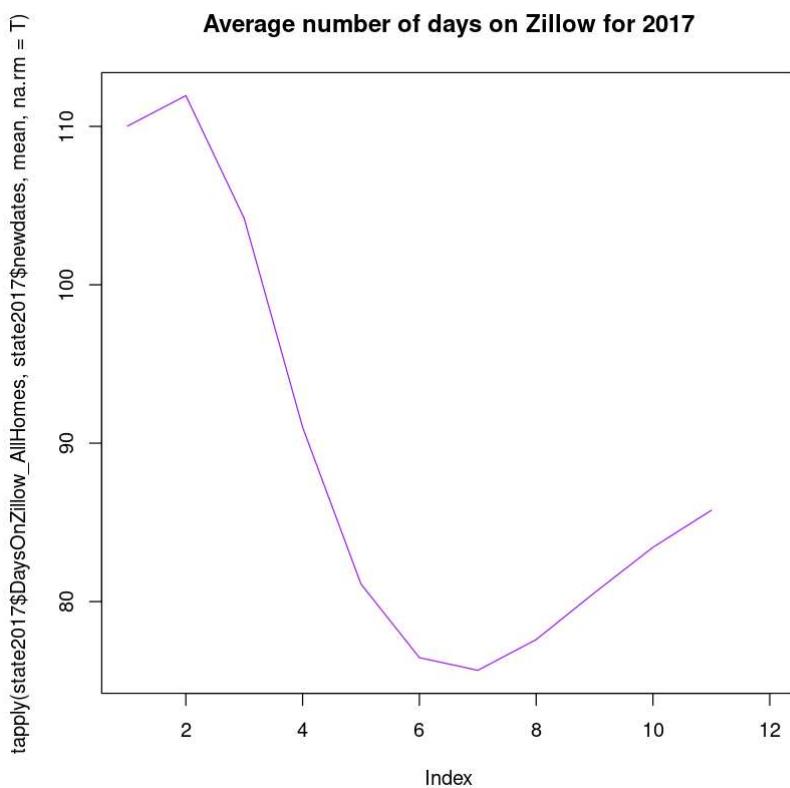
```
In [40]: tapply(state2017$DaysOnZillow_AllHomes,
            state2017$newdates, mean, na.rm=T)
# since there are lot of NA's. i will remove them
```

	2017-01-31	2017-02-28	2017-03-31	2017-04-30	2017-05-31	2017-06-30	2017-07-31
1	110.01716	111.94118	104.18627	91.00980	81.11029	76.46078	75.66176
2	2017-08-31	2017-09-30	2017-10-31	2017-11-30	2017-12-31		
	77.60294	80.56127	83.42647	85.76471			
							NaN

```
In [41]: plot(tapply(state2017$DaysOnZillow_AllHomes,
                  state2017$newdates, mean, na.rm=T))
# by the default we have dots
# default is dot if the type is not specified.
```



```
In [52]: plot(tapply(state2017$DaysOnZillow_AllHomes,
                  state2017$newdates, mean, na.rm=T),
                  type= "l", col = "purple",
                  main = "Average number of days on Zillow for 2017" )
# for Line plot, to turn it into a line
```



Find the information only for the year 2017 and call it states2017. Then create a lineplot that shows the average DaysOnZillow_AllHomes by Date using states2017 data. What do you notice? There are lot of houses during winter and fall implies that the demands are low in that period but houses are scarce during summer meaning demand will be high for houses. When was the best month/months for posting a home for sale in 2017? The best month is the month of June.

Question 5

```
In [53]: head(state$RegionName)
# the region name
[1] "Alabama"      "Arizona"       "Arkansas"      "California"    "Colorado"
[6] "Connecticut"
```

```
In [54]: table(state$RegionName)
```

Alabama	Alaska	Arizona	Arkansas
261	244	261	261
California	Colorado	Connecticut	Delaware
261	261	261	261
DistrictofColumbia	Florida	Georgia	Hawaii
233	261	261	261
Idaho	Illinois	Indiana	Iowa
261	261	261	261
Kansas	Kentucky	Louisiana	Maine
261	261	261	261
Maryland	Massachusetts	Michigan	Minnesota
261	261	261	261
Mississippi	Missouri	Montana	Nebraska
261	261	240	261
Nevada	NewHampshire	NewJersey	NewMexico
261	261	261	261
NewYork	NorthCarolina	NorthDakota	Ohio
261	261	159	261
Oklahoma	Oregon	Pennsylvania	RhodeIsland
261	261	261	261
SouthCarolina	SouthDakota	Tennessee	Texas
261	260	261	261
UnitedStates	Utah	Vermont	Virginia
95	261	260	261
Washington	WestVirginia	Wisconsin	Wyoming
261	261	261	237

```
In [63]: IndianaDF <- subset(state2017, RegionName == "Indiana" )
```

```
In [60]: MaineDF <- subset(state2017, RegionName == "Maine" )
```

```
In [61]: HawaiiDF <- subset(state2017, RegionName == "Hawaii" )
```

```
In [64]: dim(IndianaDF)
```

```
[1] 12 87
```

```
In [65]: dim(MaineDF)
```

```
[1] 12 87
```

```
In [66]: dim(HawaiiDF)
```

since i got same value of 12 rows and 87 columns for every states
very likely one row there per month

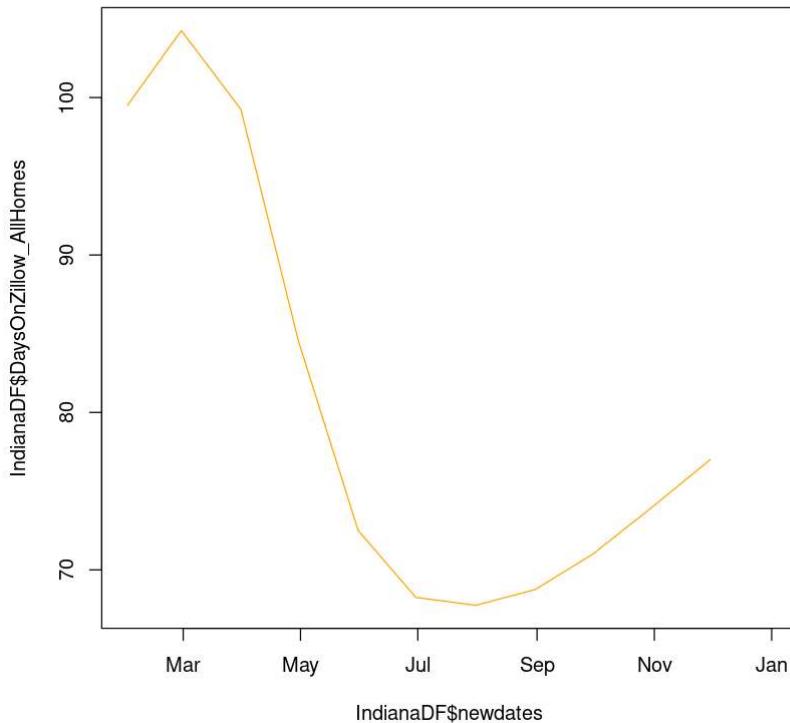
```
[1] 12 87
```

```
In [69]: plot(IndianaDF$newdates,IndianaDF$DaysOnZillow_AllHomes,
```

type = "l", col = "orange")

IndianaDF\$DaysOnZillow_ALLHomes is for y axis,

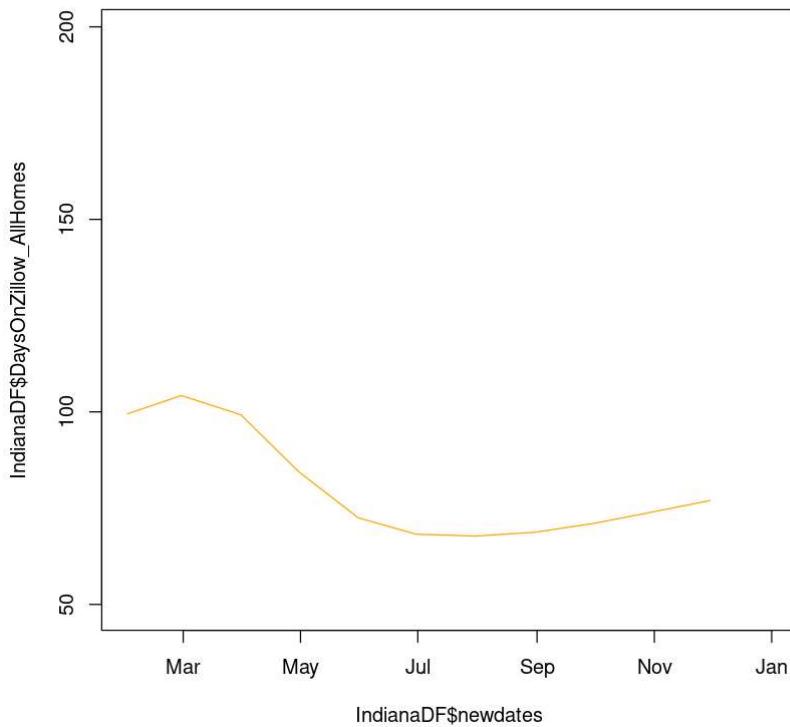
we are having several different plot, for each state, then we will put different col



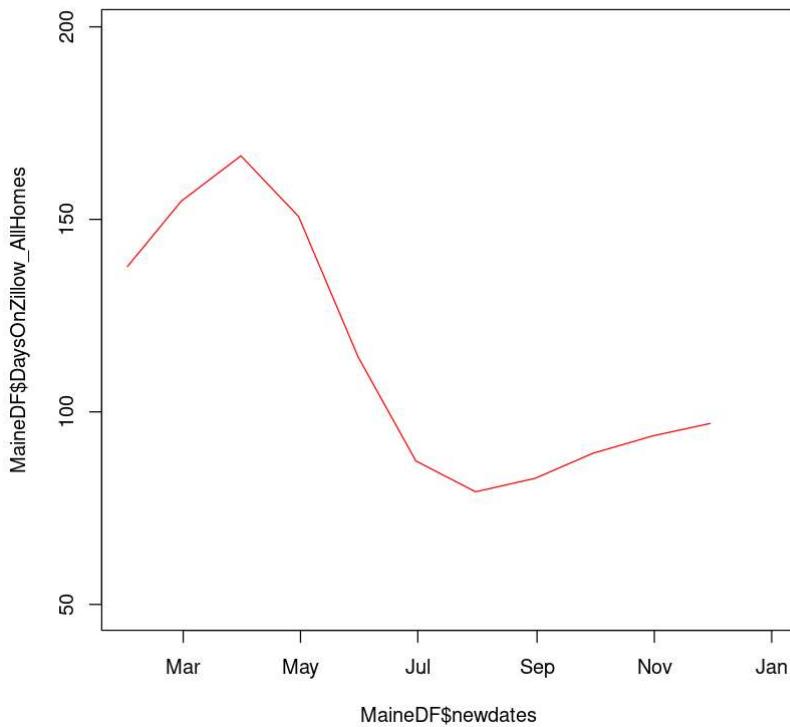
```
In [71]: min(state2017$DaysOnZillow_AllHomes, na.rm=T)
# i want to make sure that i get my ylim values correct
# this is the ylim minimum value
[1] 49.25
```

```
In [73]: max(state2017$DaysOnZillow_AllHomes, na.rm=T)
# this is the ylim maximum value
[1] 198.5
```

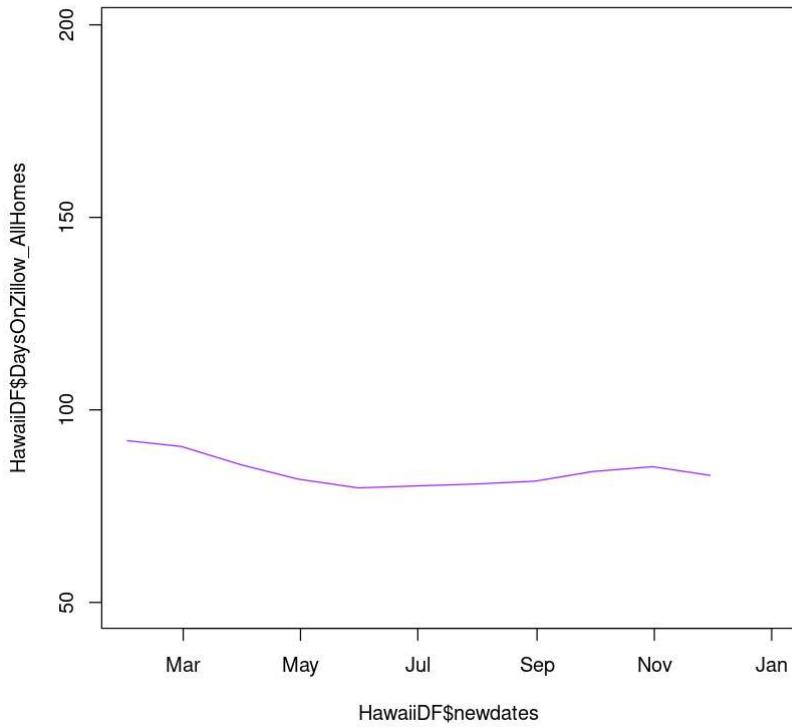
```
In [74]: plot(IndianaDF$newdates, IndianaDF$DaysOnZillow_AllHomes,
        type = "l", col = "orange",
        ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
                 max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
# new plot for Indiana after adjusting the ylim
```



```
In [81]: plot(MaineDF$newdates, MaineDF$DaysOnZillow_AllHomes,  
           type = "l", col = "red",  
           ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),  
                   max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
```

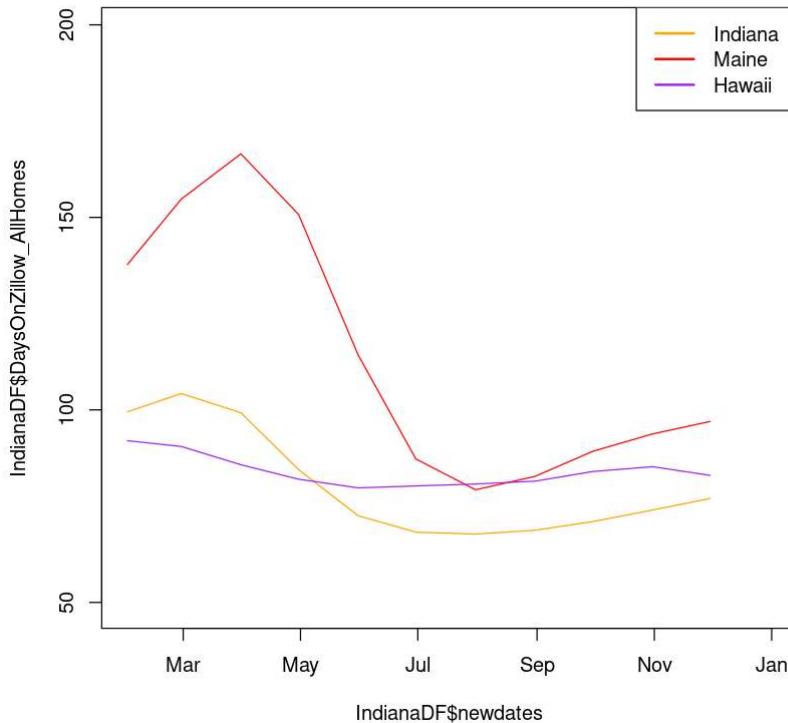


```
In [80]: plot(HawaiiDF$newdates,HawaiiDF$DaysOnZillow_AllHomes,
           type = "l", col = "purple",
           ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
                     max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
```



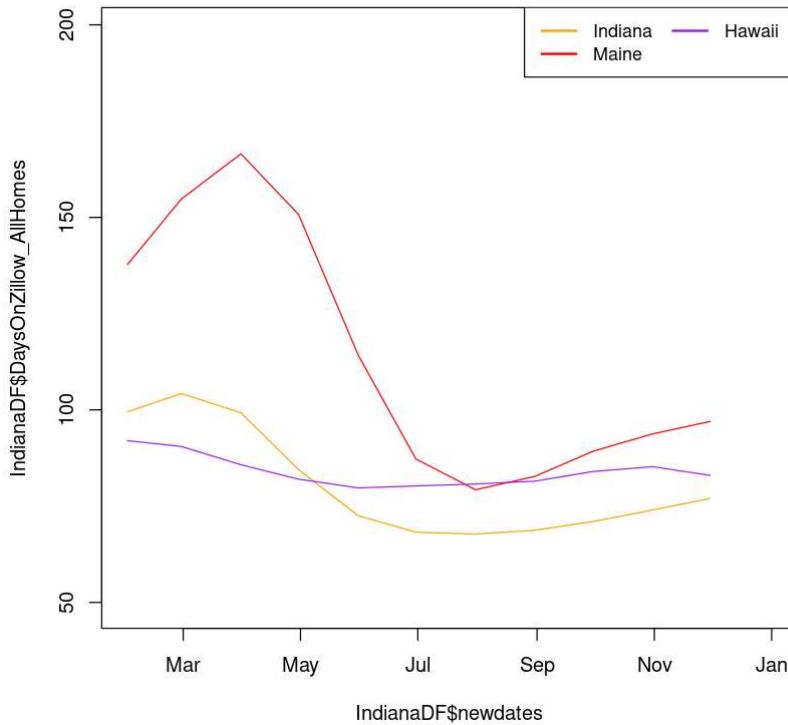
```
In [91]: plot(IndianaDF$newdates,IndianaDF$DaysOnZillow_AllHomes,
           type = "l", col = "orange",
           ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
                     max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
)
lines(MaineDF $newdates,MaineDF $DaysOnZillow_AllHomes,
      type = "l", col = "red",
      ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
                     max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
)
lines(HawaiiDF$newdates,HawaiiDF$DaysOnZillow_AllHomes,
      type = "l", col = "purple",
      ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
                     max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
)
legend( "topright", legend = c ("Indiana", "Maine", "Hawaii"),
        lwd = 2, col = c("orange", "red", "purple"))

# to have them all on the same Line, and also with Legend.
# it will plot that of plot and have the lines on it.
# my Line width (Lwd) is 2
```



```
In [103]: plot(IndianaDF$newdates, IndianaDF$DaysOnZillow_AllHomes,
      type = "l", col = "orange",
      ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
              max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
      )
lines(MaineDF $newdates,MaineDF $DaysOnZillow_AllHomes,
      type = "l", col = "red",
      ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
              max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
      )
lines(HawaiiDF$newdates,HawaiiDF$DaysOnZillow_AllHomes,
      type = "l", col = "purple",
      ylim = c(min(state2017$DaysOnZillow_AllHomes, na.rm=T),
              max(state2017$DaysOnZillow_AllHomes, na.rm=T)))
      )
legend( "topright", legend = c ("Indiana", "Maine", "Hawaii"),
       lwd = 2, col = c("orange", "red", "purple"), ncol=2, cex = .9)

# ncol = 2, means to split the column into two incase of several list
# cex = .9 means to change the size to whatever size you desire.
```



Now we want to know if homes sell faster in different states? Lets look at Indiana, Maine, and Hawaii. Create a lineplot that uses DaysOnZillow_AllHomes by Date with one line per state. Use the states2017 dataset for this question. Make sure to have each state line colored differently and have a legend to identify which is which.

Pledge

By submitting this work I hereby pledge that this is my own, personal work. I've acknowledged in the designated place at the top of this file all sources that I used to complete said work, including but not limited to: online resources, books, and electronic communications. I've noted all collaboration with fellow students and/or TA's. I did not copy or plagiarize another's work.

As a Boilermaker pursuing academic excellence, I pledge to be honest and true in all that I do. Accountable together – We are Purdue.