A7.1

I work in a financial institution. Typically most of the bill payments and money movements are around mid month, month end or on saturday. Seasonality also plays a big role in understanding money transfer transactions (like increased spending due to holidays).

Exponential smoothing would be a good model to apply to money movement transactions, as they have trends and seasonality.

In my estimate values of alpha, beta and gamma would be more closer to 1, because the variation in daily number of transactions would be due to actual changes/factors with little or no randomness.

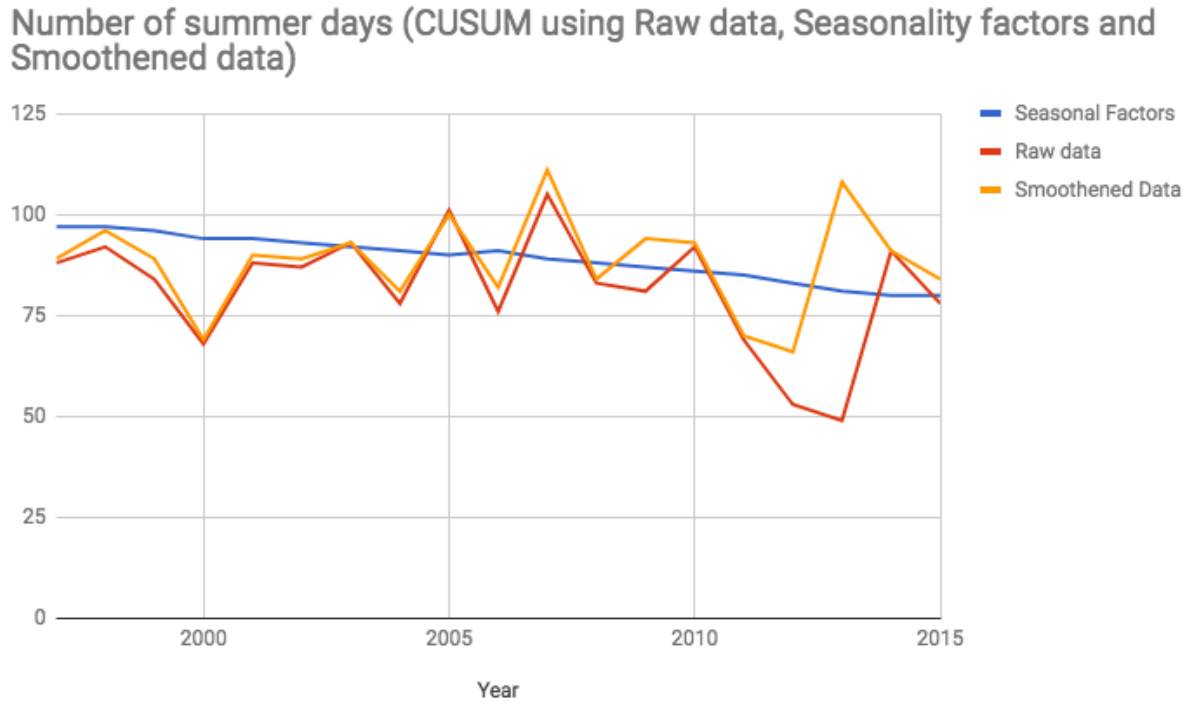
A7.2
Use HoltWinters in R to apply exponential smoothing. HoltWinters algorithm can be applied in few different ways:
1- Without trend and seasonality
2- With trend only
3- With trend and seasonality as “additive”
4- With trend and seasonality as “multiplicative”.

I have have used smoothing coefficients retrieved from the 4th option above, and applied CUSUM function using excel to identify if the summer in Atlanta has gotten later over 20 years.
(attached Excel ans7_2.xlsx)

C & T values: Using C and T based on standard deviation was not shown any change in St values, so I experimented with various C and T values. For this analysis I used C =0, and T = 4.

As listed below, I also used smoothened temperatures and applied CUSUM function.

Chart below shows the number of summer days in each year using CUSUM on Smoothened coefficients, on Smoothened data and on Raw temperature data (from week 2 solution of 6.2).



Conclusion: Based on the analysis it can be concluded that summer in Atlanta has not gotten later.

A8.1

One example where linear regression would be helpful in my organization is predicting the number of customers, over next 5 years, who would stop coming to the branch and would be using other means of banking like internet or mobile.

Predictors could be:

- Customers Age
- Type of service required
- Service available via other channels
- Customers net value with bank
- Customers Profession
- Customers Education Level

A 8.2
Using lm linear model regression to predict observed crime rate for the given data.

I created 3 models as listed below:

	Factors	R squared	R squared adjusted	R squared with cross validation	R squared adjusted with cross validation	Crime Prediction
Model 1 (using lm)	All 15	0.8031	0.7078	0.413	0.13	155.4349
Model 2 (using lm)	Only with p-value of 0.05 or less (as given by output of Model 1) **factors:** M + Ed + Po1 + U2 + Ineq + Prob	0.7659	0.7307	0.638	0.584	1304.245
Model 3 (using lm)	Using formula based model by AIC **factors:** M + Ed + Po1 + M.F + U1 + U2 +Ineq + Prob	0.7888	0.7444	0.564	0.473	1038.413

Model 1 output

```
##
##
##Call:
##  lm(formula = Crime ~ ., data = data)
##
##Residuals:
##  Min      1Q  Median      3Q     Max
##-395.74  -98.09   -6.69  112.99  512.67

##Coefficients:
##  Estimate Std. Error t value Pr(>|t|)
##(Intercept) -5.984e+03  1.628e+03  -3.675 0.000893 ***
##  M           8.783e+01  4.171e+01   2.106 0.043443 *
##  So         -3.803e+00  1.488e+02  -0.026 0.979765
##Ed           1.883e+02  6.209e+01   3.033 0.004861 **
##  Po1         1.928e+02  1.061e+02   1.817 0.078892 .
##Po2         -1.094e+02  1.175e+02  -0.931 0.358830
##LF          -6.638e+02  1.470e+03  -0.452 0.654654
##M.F          1.741e+01  2.035e+01   0.855 0.398995
```

```
##Pop         -7.330e-01  1.290e+00  -0.568 0.573845
##NW           4.204e+00  6.481e+00   0.649 0.521279
##U1          -5.827e+03  4.210e+03  -1.384 0.176238
##U2           1.678e+02  8.234e+01   2.038 0.050161 .
##Wealth       9.617e-02  1.037e-01   0.928 0.360754
##Ineq         7.067e+01  2.272e+01   3.111 0.003983 **
##  Prob       -4.855e+03  2.272e+03  -2.137 0.040627 *
##  Time       -3.479e+00  7.165e+00  -0.486 0.630708
##---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##Residual standard error: 209.1 on 31 degrees of freedom
##Multiple R-squared:  0.8031,     Adjusted R-squared:  0.7078
##F-statistic: 8.429 on 15 and 31 DF,  p-value: 3.539e-07
```

Model 2 output

```
##Call:
##  lm(formula = Crime ~ M + Ed + Po1 + U2 + Ineq + Prob, data = data)
##
##Residuals:
##  Min      1Q  Median      3Q     Max
##-470.68  -78.41  -19.68  133.12  556.23
##
##Coefficients:
##  Estimate Std. Error t value Pr(>|t|)
##(Intercept) -5040.50     899.84  -5.602 1.72e-06 ***
##  M            105.02      33.30   3.154  0.00305 **
##  Ed           196.47      44.75   4.390 8.07e-05 ***
##  Po1          115.02      13.75   8.363 2.56e-10 ***
##  U2            89.37      40.91   2.185  0.03483 *
##  Ineq          67.65      13.94   4.855 1.88e-05 ***
##  Prob       -3801.84    1528.10  -2.488  0.01711 *
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##Residual standard error: 200.7 on 40 degrees of freedom
##Multiple R-squared:  0.7659,     Adjusted R-squared:  0.7307
##F-statistic: 21.81 on 6 and 40 DF,  p-value: 3.418e-11
```

Model 3 output

```
##Call:
##lm(formula = Crime ~ M + Ed + Po1 + M.F + U1 + U2 + Ineq + Prob,
##   data = data)
##
##Residuals:
```

```
##  Min      1Q  Median      3Q     Max
##
##Coefficients:
##  Estimate Std. Error t value Pr(>|t|)
##(Intercept) -6426.10    1194.61  -5.379 4.04e-06 ***
##  M            93.32      33.50   2.786  0.00828 **
##  Ed          180.12      52.75   3.414  0.00153 **
##  Po1          102.65      15.52   6.613 8.26e-08 ***
##  M.F           22.34      13.60   1.642  0.10874
##U1         -6086.63    3339.27  -1.823  0.07622 .
##U2           187.35      72.48   2.585  0.01371 *
##  Ineq          61.33      13.96   4.394 8.63e-05 ***
##  Prob       -3796.03    1490.65  -2.547  0.01505 *
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##Residual standard error: 195.5 on 38 degrees of freedom
##Multiple R-squared:  0.7888,     Adjusted R-squared:  0.7444
##F-statistic: 17.74 on 8 and 38 DF,  p-value: 1.159e-10
```

Conclusion

Model 2 seems to be the best quality model given the values of adjusted R - squared with cross validation. Predicted Crime value for the given data point is 1304.
Factors used for prediction are: M, Ed, Po1, U2, Ineq and Prob.

M	percentage of males aged 14–24 in total state population
Ed	mean years of schooling of the population aged 25 years or over
Po1	per capita expenditure on police protection in 1960
U2	unemployment rate of urban males 35–39
Ineq	income inequality: percentage of families earning below half the median income
Prob	probability of imprisonment: ratio of number of commitments to number of offenses