ESM244 Assignment3

Craig O’Neill

March 12, 2018

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 3.3.3

## -- Attaching packages ---------------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 2.2.1.9000 v purrr 0.2.4   
## v tibble 1.4.2 v dplyr 0.7.4   
## v tidyr 0.7.2 v stringr 1.2.0   
## v readr 1.1.1 v forcats 0.2.0

## Warning: package 'tibble' was built under R version 3.3.3

## Warning: package 'tidyr' was built under R version 3.3.3

## Warning: package 'readr' was built under R version 3.3.3

## Warning: package 'purrr' was built under R version 3.3.3

## Warning: package 'dplyr' was built under R version 3.3.3

## Warning: package 'stringr' was built under R version 3.3.3

## Warning: package 'forcats' was built under R version 3.3.3

## -- Conflicts ------------------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(forecast)

## Warning: package 'forecast' was built under R version 3.3.3

##   
## Attaching package: 'forecast'

## The following object is masked from 'package:ggplot2':  
##   
## autolayer

library(tseries)

## Warning: package 'tseries' was built under R version 3.3.3

library(sf)

## Warning: package 'sf' was built under R version 3.3.3

## Linking to GEOS 3.6.1, GDAL 2.2.0, proj.4 4.9.3

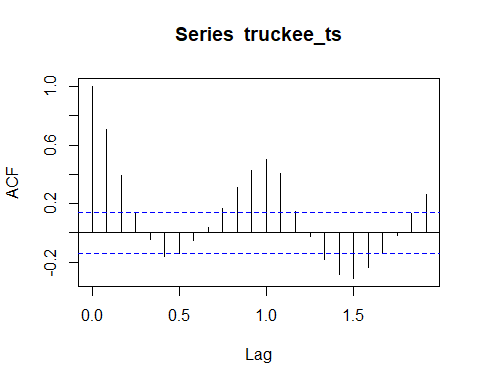
library(tmap)

## Warning: package 'tmap' was built under R version 3.3.3

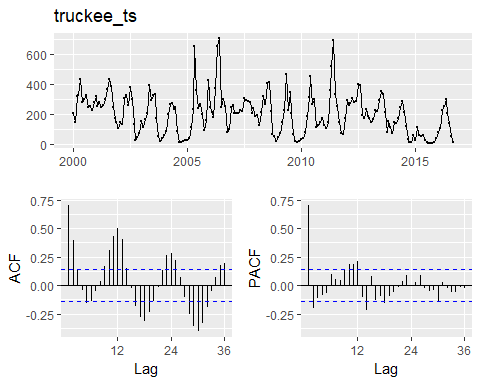
## Task 3

# a) Create a graph of the decomposed time series data (tuckee\_flow). In text beneath the graph describe the time series data.

truckee <- read.csv("truckee\_flow-edit.csv")  
  
truckee\_ts <- ts(truckee$mean\_va, frequency = 12, start = c(2000, 1))  
  
truckee\_dc <- decompose(truckee\_ts, type = "additive")  
acf(truckee\_ts)



ggtsdisplay(truckee\_ts)



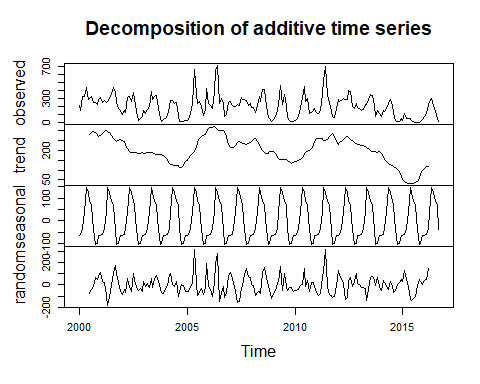
adf\_truckee <- adf.test(truckee\_ts)

## Warning in adf.test(truckee\_ts): p-value smaller than printed p-value

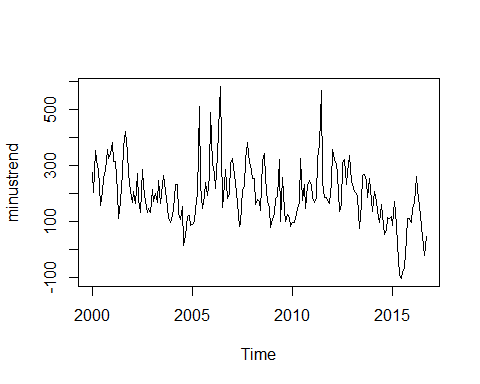
adf\_truckee

##   
## Augmented Dickey-Fuller Test  
##   
## data: truckee\_ts  
## Dickey-Fuller = -5.5387, Lag order = 5, p-value = 0.01  
## alternative hypothesis: stationary

plot(truckee\_dc)



minustrend <- truckee\_dc$x - truckee\_dc$seasonal  
plot(minustrend)

 Part a Response:

The time series data for Truckee discharge is stationary with an annual seasonality. A Dickey-Fuller test was used to test the data, a p-value of 0.01 was found indicating that the data is not non-stationary. The data appears to be additive, as confirmed by the even distribution of residuals in the decomposed dataset. There appears to be a longer term cyclical trend beyond the annual seasonality, this longer term cyclical trend appears to be occuring every 5 years. There also appears to be an outlier in the data occuring shortly after 2015.

Part a cut content: Decomposing the time series data for the Truckee discharge

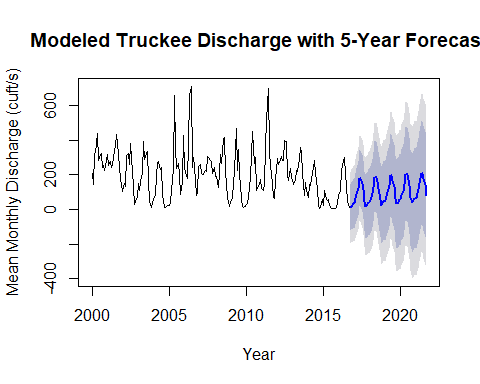
The residuals of the additive decomposed time series analysis do not seem to differ significantly throughout the time series indicating that the data is additive. The data is stationary appears to be stationary, confirmed using a Dickey-Fuller test which calculates the p-value as less than 0.05, rejecting the null-hypothesis that the data is non-stationary. Additionally, graphing the observations less the trend in the decomposed dataset reveals that the data set does not appear to have any moving trend. There is an observed outlier, however, occuring shortly after 2015.

The data appears to have annual seasonality.

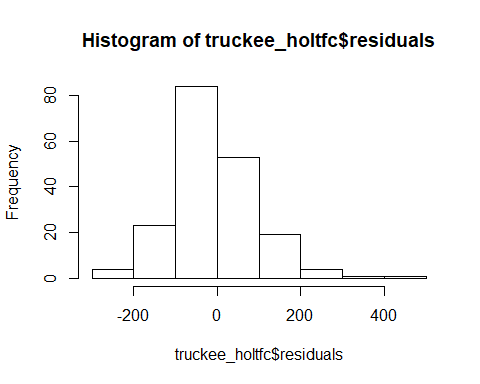
# b) Use either Holt-Winters of ARIMA to forecast Truckee River for 5 years after the final observation. Visualize the forecasted data in a graph.

First Holt-Winters

truckee\_hw <- HoltWinters(truckee\_ts)  
#hw\_fc\_60 <- predict(truckee\_hw, n.ahead = 60, prediction.interval = T, level = 0.95)  
#plot(truckee\_hw, hw\_fc\_60)  
  
truckee\_holtfc <- forecast(truckee\_hw, h = 60)  
plot(truckee\_holtfc,  
 main = "Modeled Truckee Discharge with 5-Year Forecast",  
 xlab = "Year",  
 ylab = "Mean Monthly Discharge (cuft/s)")



hist(truckee\_holtfc$residuals)



Part B) Response (Graphs above)

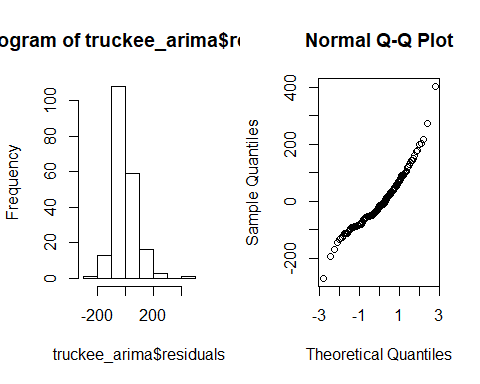
Next ARIMA

truckee\_pdq <- auto.arima(truckee\_ts)  
truckee\_pdq

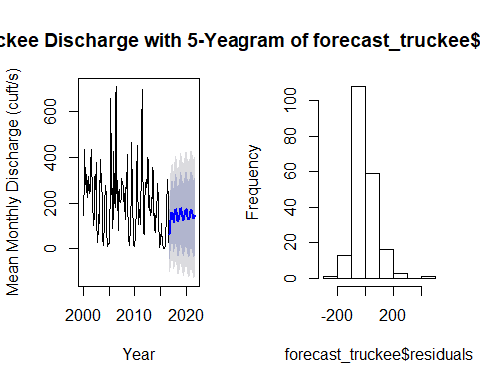
## Series: truckee\_ts   
## ARIMA(3,1,4)(0,0,1)[12]   
##   
## Coefficients:  
## ar1 ar2 ar3 ma1 ma2 ma3 ma4 sma1  
## 2.4054 -2.1585 0.6722 -2.7717 2.7951 -1.0522 0.0336 0.2230  
## s.e. 0.1122 0.1964 0.1150 0.1555 0.4059 0.3965 0.1459 0.0722  
##   
## sigma^2 estimated as 7413: log likelihood=-1173.33  
## AIC=2364.66 AICc=2365.61 BIC=2394.35

ARIMA (3,1,4)(0,0,1)

truckee\_arima <- arima(truckee\_ts, order = c(3,1,4), seasonal = list(order = c(0,0,1)))  
  
par(mfrow = c(1,2))  
hist(truckee\_arima$residuals)  
qqnorm(truckee\_arima$residuals)



forecast\_truckee <- forecast(truckee\_arima, h = 60)  
plot(forecast\_truckee,  
 main = "Modeled Truckee Discharge with 5-Year ARIMA Forecast",  
 xlab = "Year",  
 ylab = "Mean Monthly Discharge (cuft/s)")  
  
hist(forecast\_truckee$residuals)



## Task 4 Mapping California’s National Parks

ca\_shape <- st\_read(dsn = ".", layer = "california\_county\_shape\_file")

## Reading layer `california\_county\_shape\_file' from data source `C:\Users\oneil\Documents\github\esm244assgn3' using driver `ESRI Shapefile'  
## Simple feature collection with 68 features and 9 fields  
## geometry type: POLYGON  
## dimension: XY  
## bbox: xmin: -124.4096 ymin: 32.53416 xmax: -114.1344 ymax: 42.00952  
## epsg (SRID): NA  
## proj4string: NA

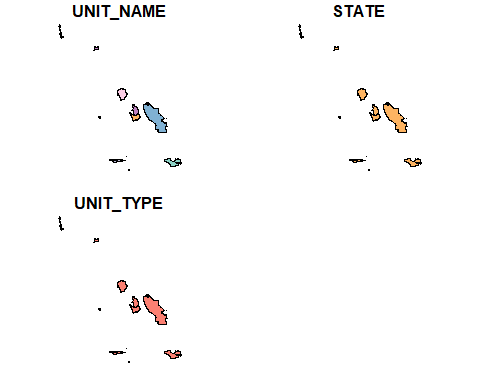
st\_crs(ca\_shape) <- 4326  
  
np\_shape <- st\_read(dsn = ".", layer = "nps\_boundary")

## Reading layer `nps\_boundary' from data source `C:\Users\oneil\Documents\github\esm244assgn3' using driver `ESRI Shapefile'  
## Simple feature collection with 419 features and 11 fields  
## geometry type: MULTIPOLYGON  
## dimension: XY  
## bbox: xmin: -170.7276 ymin: -14.28316 xmax: 145.7318 ymax: 68.65539  
## epsg (SRID): 4269  
## proj4string: +proj=longlat +datum=NAD83 +no\_defs

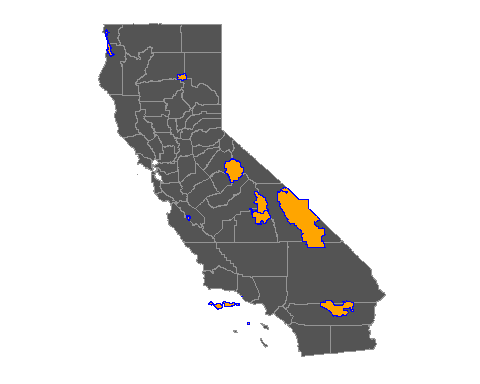
np\_shape <- st\_transform(np\_shape, "+init=epsg:4326")  
  
  
canp\_shape <- np\_shape %>%   
 filter(STATE == "CA") %>%  
 filter(UNIT\_TYPE == "National Park") %>%   
 select(UNIT\_NAME, STATE, UNIT\_TYPE, geometry)

## Warning: package 'bindrcpp' was built under R version 3.3.3

View(ca\_shape)   
plot(canp\_shape)



ca\_np\_gg <- ggplot(canp\_shape)+  
 geom\_sf(data = ca\_shape, aes(fill = STATE), color = "gray58", fill = "gray33")+  
 geom\_sf(aes(color = UNIT\_TYPE, fill = UNIT\_TYPE), color = "blue", fill = "orange")+  
 theme(axis.text.x = element\_blank(),  
 axis.text.y = element\_blank(),  
 axis.ticks = element\_blank(),  
 rect = element\_blank(),  
 panel.grid.major = element\_line(color = "transparent"))  
  
ca\_np\_gg



### Task 5 LTER Lizards in the Northern Chihuahuan Desert

library(VIM)

## Warning: package 'VIM' was built under R version 3.3.3

## Loading required package: colorspace

## Loading required package: grid

## Loading required package: data.table

## Warning: package 'data.table' was built under R version 3.3.3

##   
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':  
##   
## between, first, last

## The following object is masked from 'package:purrr':  
##   
## transpose

## VIM is ready to use.   
## Since version 4.0.0 the GUI is in its own package VIMGUI.  
##   
## Please use the package to use the new (and old) GUI.

## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues

##   
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':  
##   
## sleep

library(mice)

## Warning: package 'mice' was built under R version 3.3.3

## Loading required package: lattice

##   
## Attaching package: 'mice'

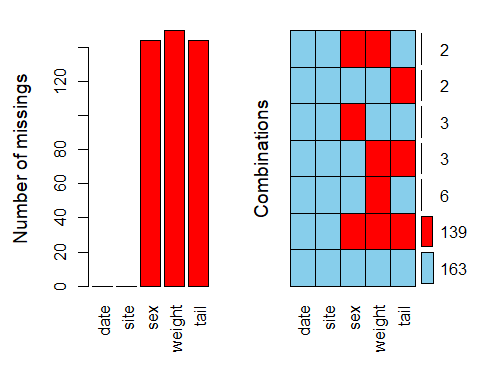
## The following object is masked from 'package:tidyr':  
##   
## complete

1. For all lizards at site “CALI”, do weights of male and female ault lizards differ significantly?

lizard <- read.csv("lter\_lizard\_pitfall\_edited.csv")  
  
lizard\_cal <- lizard %>%   
 filter(site == "CALI") %>%   
 select(date, site, sex, weight, tail)  
  
lizard\_cal$site <- as.factor(lizard\_cal$site)  
lizard\_cal$tail <- as.factor(lizard\_cal$tail)  
lizard\_cal$sex <- as.factor(lizard\_cal$sex)

Identify missing cases

liz\_cc <- lizard\_cal[complete.cases(lizard\_cal),] #this extra comma tells R to just look at the rows of data.  
liz\_missings <- lizard\_cal[!complete.cases(lizard\_cal),]  
  
aggr(lizard\_cal, prop = F, numbers = T)



imputed\_liz <- mice(lizard\_cal, m = 5)

##   
## iter imp variable  
## 1 1 sex weight tail  
## 1 2 sex weight tail  
## 1 3 sex weight tail  
## 1 4 sex weight tail  
## 1 5 sex weight tail  
## 2 1 sex weight tail  
## 2 2 sex weight tail  
## 2 3 sex weight tail  
## 2 4 sex weight tail  
## 2 5 sex weight tail  
## 3 1 sex weight tail  
## 3 2 sex weight tail  
## 3 3 sex weight tail  
## 3 4 sex weight tail  
## 3 5 sex weight tail  
## 4 1 sex weight tail  
## 4 2 sex weight tail  
## 4 3 sex weight tail  
## 4 4 sex weight tail  
## 4 5 sex weight tail  
## 5 1 sex weight tail  
## 5 2 sex weight tail  
## 5 3 sex weight tail  
## 5 4 sex weight tail  
## 5 5 sex weight tail

summary(imputed\_liz)

## Multiply imputed data set  
## Call:  
## mice(data = lizard\_cal, m = 5)  
## Number of multiple imputations: 5  
## Missing cells per column:  
## date site sex weight tail   
## 0 0 144 150 144   
## Imputation methods:  
## date site sex weight tail   
## "" "" "polyreg" "pmm" "logreg"   
## VisitSequence:  
## sex weight tail   
## 3 4 5   
## PredictorMatrix:  
## date site sex weight tail  
## date 0 0 0 0 0  
## site 0 0 0 0 0  
## sex 1 0 0 1 1  
## weight 1 0 1 0 1  
## tail 1 0 1 1 0  
## Random generator seed value: NA

COmpete the cases using imputed cvalues

com\_liz <- complete(imputed\_liz, action = 2)

lm\_liz <- with(imputed\_liz, lm(weight ~ sex))  
impute\_pooled\_liz <- pool(lm\_liz)  
  
impute\_pooled\_liz

## Call: pool(object = lm\_liz)  
##   
## Pooled coefficients:  
## (Intercept) sex2 sex3   
## 5.01406714 -3.06201244 -0.02437578   
##   
## Fraction of information about the coefficients missing due to nonresponse:   
## (Intercept) sex2 sex3   
## 0.9484200 0.8903486 0.7939250

summary(impute\_pooled\_liz)

## est se t df Pr(>|t|)  
## (Intercept) 5.01406714 1.444259 3.47172328 3.863709 0.02700501  
## sex2 -3.06201244 1.723743 -1.77637381 4.907613 0.13692935  
## sex3 -0.02437578 1.213531 -0.02008666 6.696027 0.98455961  
## lo 95 hi 95 nmis fmi lambda  
## (Intercept) 0.9477564 9.080378 NA 0.9484200 0.9272099  
## sex2 -7.5182428 1.394218 NA 0.8903486 0.8532265  
## sex3 -2.9205396 2.871788 NA 0.7939250 0.7403713

#combined\_liz <- complete(impute\_pooled\_liz)

1. For All Lizards trapped at site CALI do weights of male and female adult lizards

liz\_cc\_f <- liz\_cc %>%   
 filter(sex == "F")  
liz\_cc\_m <- liz\_cc %>%   
 filter(sex == "M")  
  
  
t.test(liz\_cc\_f$weight, liz\_cc\_m$weight)

##   
## Welch Two Sample t-test  
##   
## data: liz\_cc\_f$weight and liz\_cc\_m$weight  
## t = 0.6825, df = 123.53, p-value = 0.4962  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.398063 2.869579  
## sample estimates:  
## mean of x mean of y   
## 5.826667 5.090909

liz\_cc\_f <- com\_liz %>%   
 filter(sex == "F")  
liz\_cc\_m <- com\_liz %>%   
 filter(sex == "M")  
  
  
t.test(liz\_cc\_f$weight, liz\_cc\_m$weight)

##   
## Welch Two Sample t-test  
##   
## data: liz\_cc\_f$weight and liz\_cc\_m$weight  
## t = 1.0333, df = 151.81, p-value = 0.3031  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.8576659 2.7383477  
## sample estimates:  
## mean of x mean of y   
## 5.864583 4.924242

Task 5 A)

The mean weight of lizards trapped at the CALI site do signficantly differ between males and females (Welch Two Sample t-test, p-value = 0.0142, alpha = 0.05). Missing values in dataset were imputed using multiple imputation (m=5).

1. For lizards trapped at CALI site, is there a significant difference in the proportion of adult male and female lizards with broken tails?

liz\_cc\_con <- com\_liz %>%  
 select(sex, tail) %>%   
 filter(sex != "J")  
  
table1 <- table(liz\_cc\_con)  
table2 <- table1[-2,]  
  
View(table2)  
  
chisq <- chisq.test(table2)  
chisq

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: table2  
## X-squared = 0.50259, df = 1, p-value = 0.4784

Task 5 1)

A chi-square analysis revealed significant difference between proportions of lizards with broken tails based upon sex (X2(1) = 10.353, p = 0.001293) trapped at the CALI site. Multiple imputation was used (m=5) to address missing at random data.