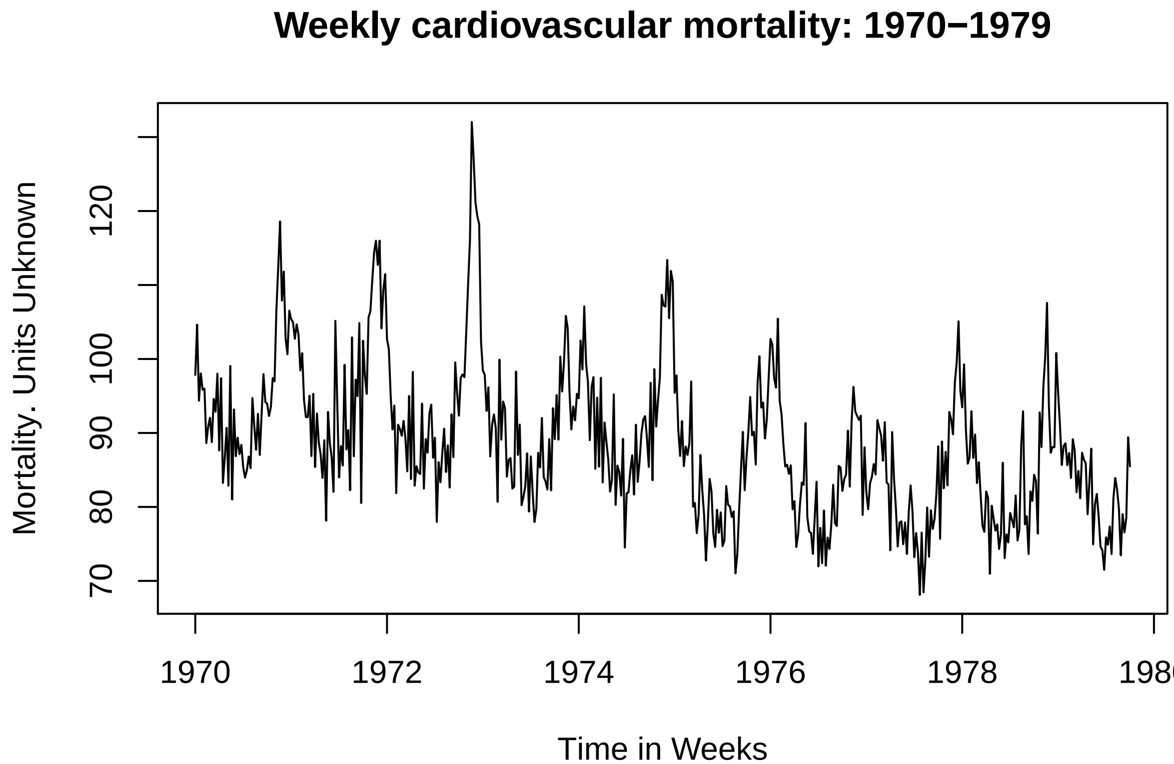
Krysten Thompson w271: Homework 5

Professor Jeffrey Yau

```r
rm(list = ls())
library(astsa)

plot(cmort, xlab= "Time in Weeks", ylab="Mortality. Units Unknown")
title(main="Weekly cardiovascular mortality: 1970-1979")
```



1. Conduct the EDA of the `weekly cardiovascular mortality` time series.

```r
cmort
```

```
## Time Series:
## Start = c(1970, 1)
## End = c(1979, 40)
## Frequency = 52
##   [1]  97.85 104.64  94.36  98.05  95.85  95.98  88.63  90.85  92.06  88.75
##  [11]  94.60  92.86  98.02  87.64  97.40  83.24  86.60  90.69  82.86  99.06
##  [21]  81.00  93.18  86.86  89.35  87.13  88.39  85.38  83.96  84.95  86.81
##  [31]  85.25  94.72  90.96  87.73  92.58  87.02  92.68  97.95  94.19  93.95
##  [41]  92.29  93.48  97.41  96.98 106.60 112.41 118.59 107.90 111.82 102.78
##  [51] 100.63 106.54 105.39 104.99 102.72 104.67 103.30  98.46 100.78  94.43
##  [61]  92.16  92.13  95.04  86.88  95.29  85.39  92.63  88.72  87.09  83.92
##  [71]  89.01  78.14  92.84  88.65  86.66  82.04 105.15  94.19  83.99  88.22
```

```
##  [81]  85.59  99.21  87.77  90.39  82.25 102.92  86.84  97.22  94.97 104.86
##  [91]  80.53 102.47  97.80  95.28 105.64 106.44 110.65 114.41 115.98 112.72
## [101] 115.99 104.14 109.30 111.47 102.67 101.34  94.97  90.48  93.70  81.87
## [111]  91.12  90.55  89.62  91.63  89.22  84.81  95.00  83.77  98.26  82.87
## [121]  85.51  84.67  84.45  93.96  82.47  89.18  87.33  92.67  93.83  86.65
## [131]  89.35  77.97  86.06  83.34  87.35  90.58  84.72  88.32  82.62  92.52
## [141]  86.73  99.54  95.50  92.34  97.49  97.90  97.54 103.78 110.17 116.04
## [151] 132.04 126.95 121.11 119.30 118.20 102.36  98.44  97.87  92.98  96.16
## [161]  86.83  91.09  92.50  90.64  80.69  99.91  89.09  94.25  93.33  84.08
## [171]  86.37  86.61  82.47  82.73  98.29  87.02  91.13  80.23  81.36  82.80
## [181]  87.24  79.37  86.83  82.00  77.96  79.69  87.32  85.33  92.02  84.03
## [191]  83.45  82.35  89.17  82.24  93.35  89.16  95.13  89.10 100.32  95.61
## [201]  99.61 105.82 104.20  95.45  90.48  93.58  91.71  95.29  94.69 102.50
## [211]  98.58 107.12  99.29  96.85  89.02  96.27  97.55  85.16  94.77  85.45
## [221]  97.46  83.31  91.41  88.84  86.44  82.08  83.63  95.22  80.28  85.59
## [231]  84.68  81.53  89.20  74.51  81.85  81.97  85.01  86.98  81.67  91.11
## [241]  83.38  86.18  89.93  91.82  92.27  89.29  85.41  96.78  83.60  98.63
## [251]  90.86  94.36  97.52 108.68 107.23 107.08 113.39 105.51 111.90 110.44
## [261]  95.42  97.78  90.38  86.89  91.59  85.51  88.16  87.03  88.61  96.96
## [271]  80.03  80.53  76.46  78.85  87.03  82.27  78.92  72.75  77.75  83.79
## [281]  82.13  76.42  74.58  79.61  76.52  79.27  74.70  75.49  82.82  80.29
## [291]  80.05  78.64  79.40  71.02  73.55  79.60  85.08  90.14  82.26  86.75
## [301]  90.11  94.86  89.68  90.16  85.73  96.82 100.37  93.43  94.12  89.26
## [311]  91.81  97.40 102.71 101.94  97.40  96.12 105.45  94.34  92.51  88.47
## [321]  85.47  85.70  84.48  85.63  79.67  80.78  74.58  76.32  80.37  83.30
## [331]  82.98  91.34  78.57  76.68  76.47  73.66  78.61  83.40  71.96  77.17
## [341]  72.38  79.52  72.05  75.86  74.32  77.70  82.99  77.79  77.42  85.55
## [351]  85.33  82.16  83.74  84.40  90.32  82.74  91.50  96.23  92.90  92.30
## [361]  91.76  92.34  78.91  88.06  81.91  79.68  83.20  84.05  85.80  84.39
## [371]  91.74  90.65  89.67  86.24  91.46  83.30  83.05  74.14  90.12  83.84
## [381]  79.94  74.65  77.89  78.06  74.96  77.90  73.65  79.32  82.91  79.48
## [391]  73.21  76.46  73.76  68.11  76.53  68.46  72.84  79.92  73.26  79.55
## [401]  77.02  78.39  81.86  88.20  75.69  88.85  82.52  87.44  82.91  92.84
## [411]  91.85  89.82  96.84  99.56 105.10  95.70  93.46  99.27  90.39  85.85
## [421]  86.69  92.94  86.63  89.78  83.22  86.04  81.53  77.46  76.63  82.09
## [431]  81.22  70.96  80.17  78.48  76.80  77.63  74.31  76.30  85.97  73.07
## [441]  76.29  75.21  79.18  78.25  77.24  81.55  75.46  76.86  88.43  92.93
## [451]  77.65  78.74  73.63  82.11  80.81  84.34  83.48  76.40  92.77  88.09
## [461]  96.20 100.28 107.58  93.73  87.33  88.11  88.07 100.81  95.52  91.14
## [471]  85.67  88.24  88.61  85.64  87.29  83.92  89.15  87.61  81.99  84.86
## [481]  81.17  87.33  86.40  85.90  79.01  83.49  87.88  74.94  80.32  81.75
## [491]  78.68  74.62  74.16  71.50  75.89  74.89  77.36  73.63  81.17  83.91
## [501]  82.36  79.74  73.46  79.03  76.56  78.52  89.43  85.49
```

```r
class(cmort)
```

```
## [1] "ts"
```

```r
colnames(cmort)
```

```
## NULL
```

```r
start(cmort); end(cmort); frequency(cmort)
```

```
## [1] 1970    1
```

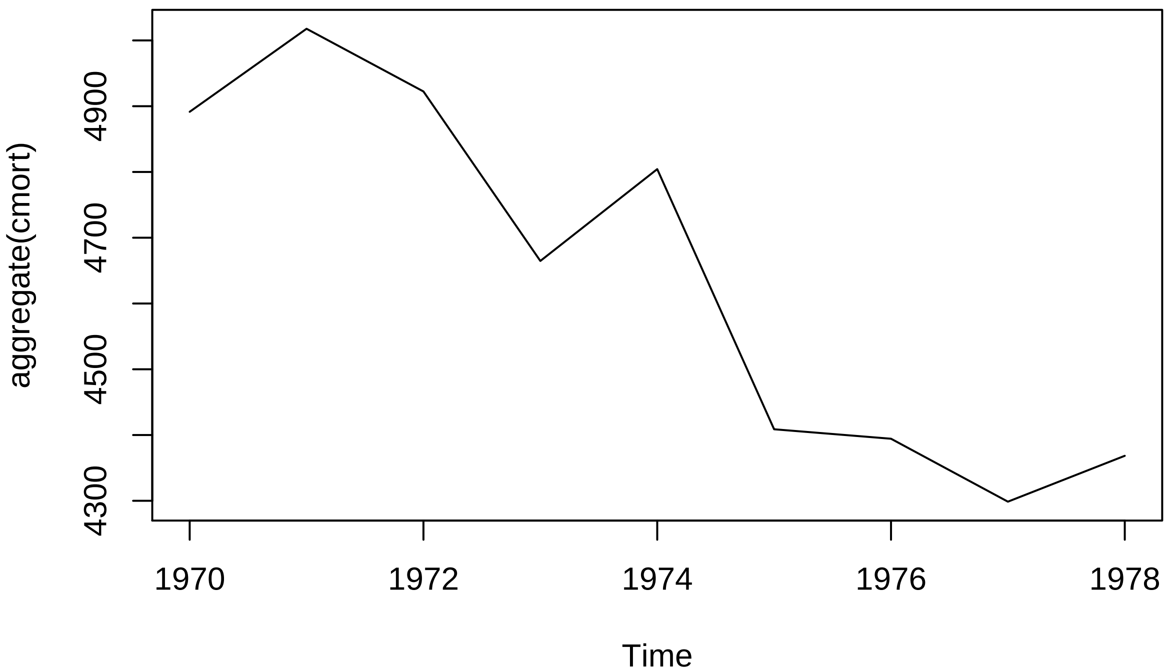
```
## [1] 1979   40
```

```
## [1] 52
```

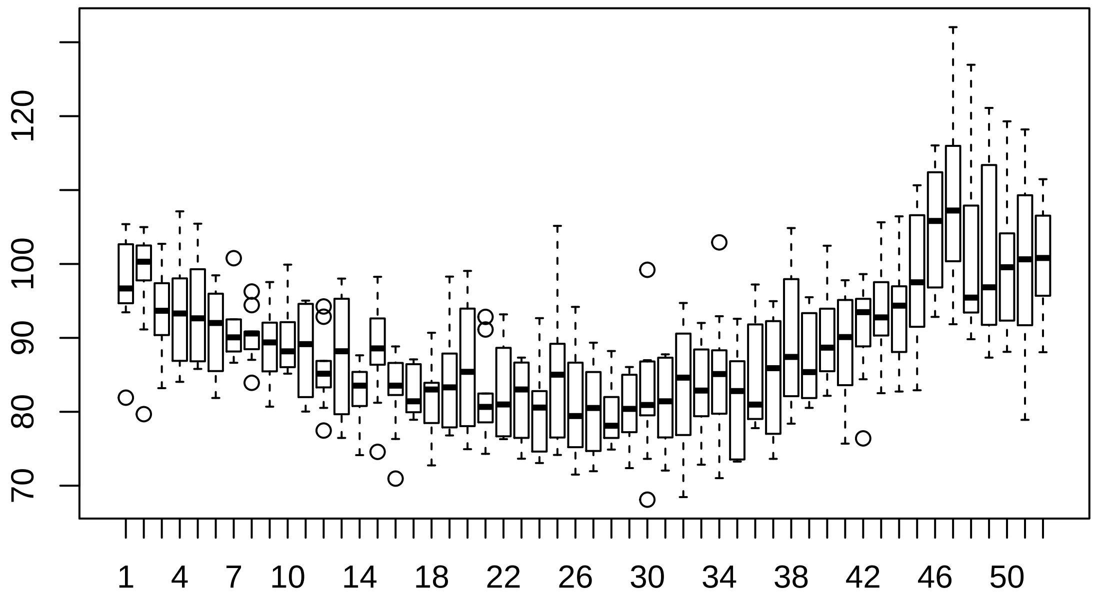
```r
summary(cmort)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   68.11   81.90   87.33   88.70   94.36  132.04
```

```r
plot(aggregate(cmort))
```

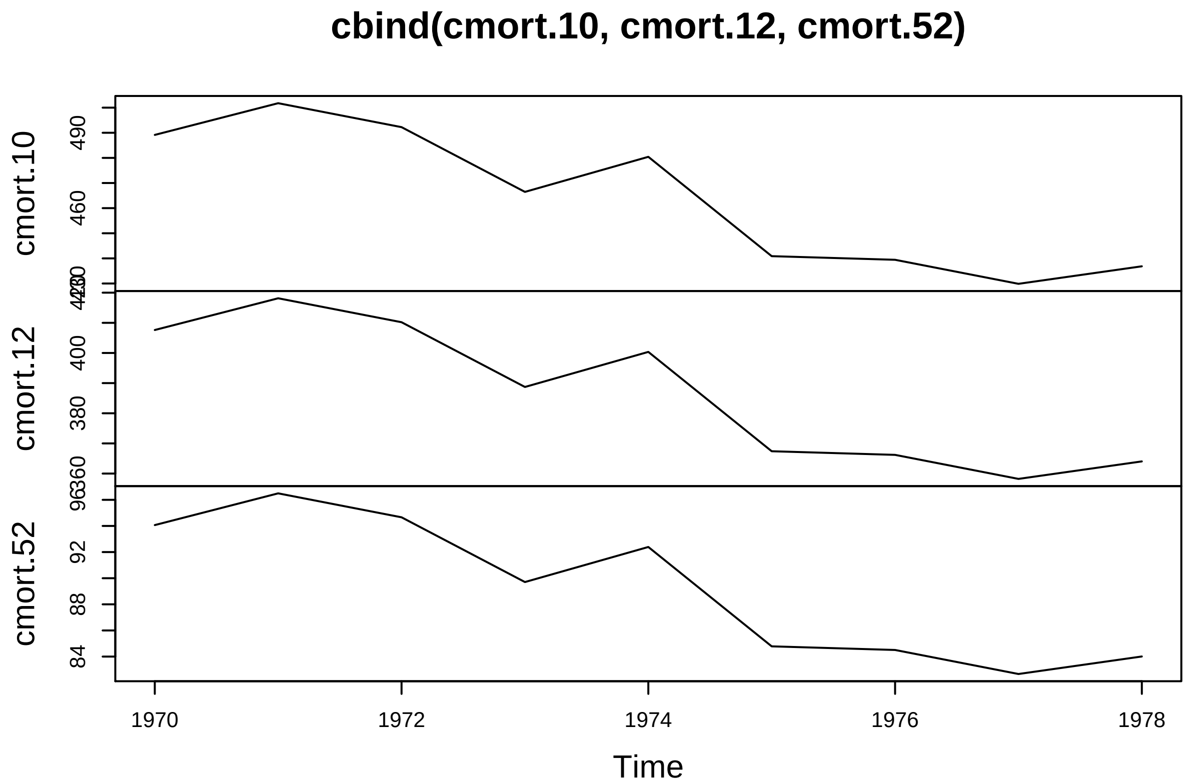



```r
boxplot(cmort ~ cycle(cmort))
```



```r
#Trying different things to explore how this all works

cmort.10 <- aggregate(cmort)/10
cmort.12 <- aggregate(cmort)/12
cmort.52 <- aggregate(cmort)/52
```

```r
plot(cbind(cmort.10, cmort.12, cmort.52))
```



```r
#I see why this doesn't work; left it in to reflect my attempts at EDA
```

```r
#want to create snapshot of 1970 to calc ratio
year.1970 <- window(cmort, start = c(1970, 1), end = c(1970, 12))

ratio.1970 <- mean(year.1970) / mean(cmort)

ratio.1970
```

```
## [1] 1.065853
```

```r
#Have to see what happened in 1972 (spike)

year.1972 <- window(cmort, start = c(1972, 1), end = c(1972, 12))

ratio.1972 <- mean(year.1972) / mean(cmort)

ratio.1972
```

```
## [1] 1.035319
```

```r
#now I'm going to create ratio snapshots at various points in 10 years

year.1973 <- window(cmort, start = c(1973, 1), end = c(1973, 12))

ratio.1973 <- mean(year.1973) / mean(cmort)

ratio.1973
```

```
## [1] 1.043277
```

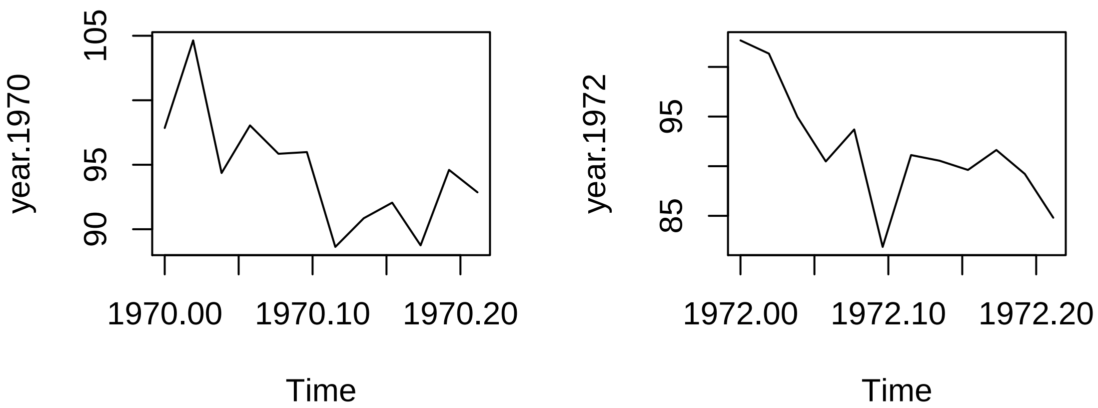
```r
#now I'm going to create ratio snapshots at various points in 10 years

year.1974 <- window(cmort, start = c(1974, 1), end = c(1974, 12))

ratio.1974 <- mean(year.1974) / mean(cmort)

ratio.1974
```

```
## [1] 1.077851
```

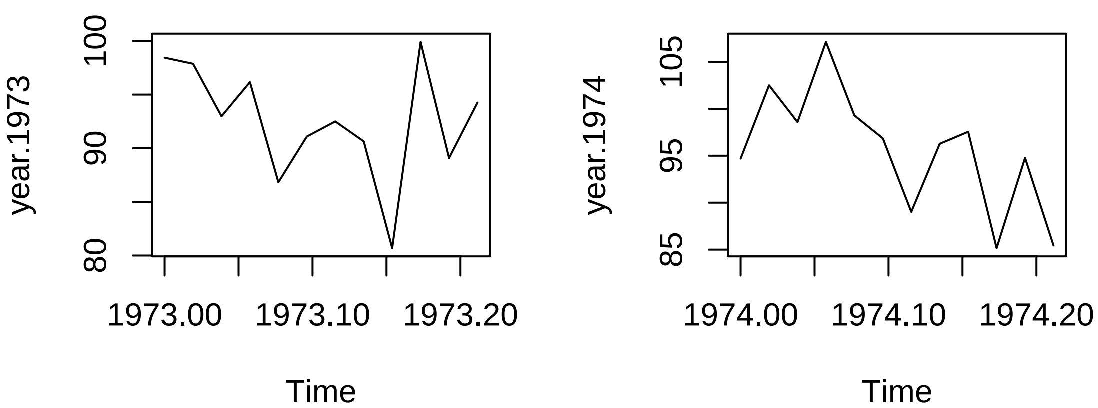
```r
#now I'm going to create ratio snapshots at various points in 10 years

year.1979 <- window(cmort, start = c(1979, 1), end = c(1979, 12))

ratio.1979 <- mean(year.1979) / mean(cmort)

ratio.1979
```

```
## [1] 0.9861455
```

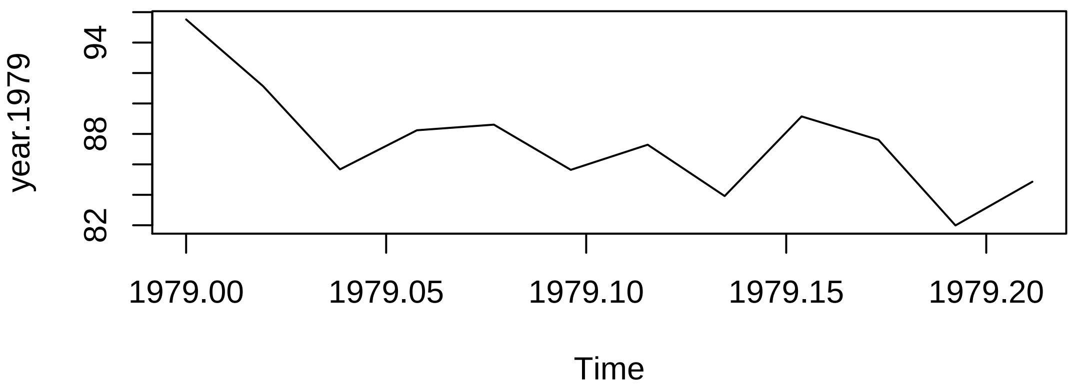
```r
par(mfrow=c(1,2))

plot(year.1970)
plot(year.1972)
```




```r
par(mfrow=c(1,2))

plot(year.1973)
plot(year.1974)
```




```r
plot(year.1979)
```



2. What features do you notice of the `weekly cardiovascular mortality` time-series plot?

There appear to be peaks and valleys which leads one to wonder if some sort of

seasonality is at play. However, I'm not sure how seasonality would effect mortality. Also, the overall slope is trending down with time.

3. Do you think that it is stationary in the mean? In the variance?

The mortality data is not stationary in the mean because the mean decreases with time. The variance appears to be stationary (with an exception in what looks to be late 1972) as the spread is generally in range of 40.

4. What pieces of information did you use from your EDA to arrive at your conclusion?

I plotted annual trends for 1970, 1972, 1973, 1974, and 1979. I also calculated the ratio of deaths in 1970, 1974, and 1979.

5. Do you find any evidence that there is a dependency structure in this time series data? Please explain.

For each of the 5 years I looked at monthly data, it appears there's a general downward trend throughout the year. Each year starts higher than it ends. But I don't know what's happening to cause this.

6. What is the difference between strict and weak stationarity?

Strict stationarity means that the probability distribution of a random variable (or joint probability distribution if more than one random variable) remains constant over time. There is no change no matter what snapshot in time is take.

Weak stationarity means that only the mean, correlation, and covariance of a random variable are invariant to the movement of time. A weak stationarity model is easier to fit because all that's needed is the mean (expected value).

7. What is the difference between an acf and pacf plot?

An acf plot starts with lag 0 while a pacf plot starts at lag 1.

8. (Open-ended question) Give two examples of questions people in your industry might ask that, based on what you learn in the async lecture, you think can be addressed using time-series analysis.

When I worked in sales for internet security products sold to consumers, I was accountable for forecasting units sold on weekly basis, as well as sales numbers. I wish I knew then what I'm learning now!