Homework 7: Solutions

36-350, Fall 2011

1. Solution

   It's difficult to make out the trends of individual states in the first plot, but it does show us that states tend to follow similar patterns. In the second plot, we can now make out the trends for individual states. Notice how some states (WY, AK, ED, ME, MT) have highly variable seasonal patterns. We can see the effects of the recent economic depression in most states, but some (AK, SD, NE, ND, IA) were barely affected.

2. Solution

```
df <- subset(urn, state == 'TX')
qplot(x = date, y = unemployment, data = df, geom = 'line')
```

   Looking at the plot closely, we see spikes occurring about every 6 months, one at the turn of each year and one in the middle of each year.

3. Solution

```
df <- ddply(df, .(year), transform, unemployment.deyeared =
            unemployment - mean(unemployment))
```

4. Solution

   We see a large spread in unemployment in few months surrounding January 1^{st}, with much less spread in the middle of the year. We also notice a large spike at June 1^{st}. We see a large spread around Jan. 1 and Dec. 1 due to the fact that we're demeaning each year, and if the unemployment rate tends to decline or increase over the course of a year then the beginning and end of the year are going to have the largest deviations from the mean.

5. Solution

```
df <- ddply(df, .(year), transform, unemployment.res =
  resid(lm(unemployment ~ month)), unemployment.fit =
  fitted(lm(unemployment ~ month)))
qplot(x = month, y = unemployment.fit, data = df, geom = 'line', group = year,
      color = year)
```

When looking at the plot of yearly linear fits, one thing to notice is that most of the fits have a negative slope, with only about 4 years having a positive slope.

6. Solution

We plot the newly adjusted unemployment rates with

```
qplot(x = month, y = unemployment.res, data = df, geom = 'line', group = year,
      color = year)
```

We notice that in comparison to the residual trends from Problem 4, the same overall pattern still exists but there is much less variation at the beginning and end of the year. The large spike at June 1st is still there, but now we have a large spike at January 1st where before it was unclear what the pattern was around that time due to its large spread. Next we plot the average for each month:

```
x <- ddply(df, .(month), summarize, unemployment.monthlyeffect =
  mean(unemployment.res))
qplot(x = month, y = unemployment.monthlyeffect, data = x, geom = 'line')
```

7. Solution

```
estimateSeasonality <- function(df) {
    df <- ddply(df, .(year), transform,
                unemployment.res = resid(lm(unemployment ~ month)),
                unemployment.fit = fitted(lm(unemployment ~ month)))
    ddply(df, .(month), summarize,
          unemployment.monthlyeffect = mean(unemployment.res))
}
```

8. Solution

We use the function to plot the de-trended yearly mean for Wyoming:

```
qplot(x = month, y = unemployment.monthlyeffect,
      data = estimateSeasonality(subset(urn, state == 'WY')),
      geom = 'line')
```

The plot shows that Wyoming has a much different pattern: it has higher relative unemployment rates in the winter months with low rates in the summer months, and no spike on June 1st.

9. Solution

To apply our function to all 50 states we again use `ddply`:

```
df8 <- ddply(urn, .(state), estimateSeasonality)
```

and we form the same two plots as in Problem 1:

```
qplot(x = month, y = unemployment.monthlyeffect, data = df8, geom = 'line',
      group = state, color = state)
qplot(x = month, y = unemployment.monthlyeffect, data = df8, geom = 'line',
      group = state, facets = ~ state)
```

There does seem to be high variation among the states in certain months, with June, July, August and December having the largest variation and May and October having very little. From the second plot we can see that there are a few patterns common to most of the states, but a few states have a pattern that they share with only a few (i.e. Wyoming and Idaho, or Florida and Arizona).