The purpose of this section is to provide a numerical example for the Stable Unit Value Treatment Assumption (SUTVA) and to present an example of loess regression using `ggplot2`.

SUTVA

The very simple but fundamental assumption of stable unit treatment values can be illustrated with a numerical example. The notation is fairly simple, but the numerical example made the concept much clearer to me. Let $\mathbf{D}$ be an $N \times 1$ column of treatment values, so that D_i is an indicator of treatment for observation $i = 1, 2, \ldots, N$. Let $\mathbf{D}'$ be another vector which defines a different treatment regime. Formally, SUTVA states that if $D_i = D_i'$, then $Y_i(\mathbf{D}) = Y_i(\mathbf{D}')$. Less formally, SUTVA states that there is no spillover effect from the treatment. As stated in class, if SUTVA does not hold, then there is not a singular treatment effect, but rather the potential for an effect of a change in D_i on all observations $j = 1, 2, \ldots, N$, not just on the own observation effect. Consider the specific treatment vector $\mathbf{D}$:

```
(D <- c(1,0,1,0,0,0,0,0,0,1))
```

```
 [1] 1 0 1 0 0 0 0 0 0 1
```

Define the treatment effect to be determined by the random vector $\mathbf{x}$. This, if there were no spillover at all, $Y_i(1) = x_i$ and $Y_i(0) = 0$. However, we induce spillover by defining the treatment effect to depend on a rolling function. Specifically, define $Y_i(D_i) = \max\{D_{i+1}, D_i\} \cdot x_i$. The following applies this rule, and then prints the value of the treatment for $i = 7$.

```
library(zoo)
x <- runif(10)
Y <- rollmax(D, 2, fill = 1) * x
print(Y[7])
```

```
[1] 0
```

Now suppose we adjust the treatment regime by setting $D_8 = 1$. Under both regimes, $D_7 = D_7' = 0$, such that if SUTVA were to hold, we should observe $Y_7(\mathbf{D}) = Y_7(\mathbf{D}')$. This is not the case, however, since the value of D_8 informs the treatment effect of the seventh observation.

```
D[8] <- 1
Y <- rollmax(D, 2, fill = 1) * x
print(Y[7])
```

```
[1] 0.7799037
```

When the treatment effect of one observation depends on the treatment of another observation, then SUTVA does not hold. We are unable to estimate the treatment effect relative to the "no intervention" scenario. This is a relatively simple example, but it's conceivable that the effect across observations may be much more complicated.

Poorly defined treatments

Suppose that the value of the treatment is a function of the level of the outcome variable. This may seem a little circular, but consider a situation where the market for the outcome adjusts based on the total activity in that market. What if the individual treatment effect is a function of the level of the aggregate outcome? The treatment is poorly defined, and it may be difficult to settle on a coneptual idea of the treatment effect, let alone an empirical assessment. Suppose for example, that the treatment effect for each individual is

$x_i \cdot (\sum_i Y_i)^{-1}$. Ignore the timing of the treatment for now, and assume that the iterative adjustment process is condensed into a single time period. The higher aggregate treatment effect will induce a shift of the market supply, perhaps, and lower the individual treatment effect — a sort of congestion effect.

The following example illustrates this point. We can calculate the treatment effects for each of the ten individuals, which is a function of the aggregate treatment effect.

```
D <- c(1,0,1,0,1,0,1,1,0,1)
x <- runif(10)
(Y <- D * x * 1/sum(D * x))
```

```
 [1] 0.20911598 0.00000000 0.15488145 0.00000000 0.08860079 0.00000000
 [7] 0.05381528 0.22953690 0.00000000 0.26404959
```

If the market responds to the aggregate level of treatment — which it does by construction — then by changing the treatment regime so that observation $i = 2$ now receives treatment, the effect will be felt by all other individuals. Note that the treatment effects for the other observations are slightly lower with the restricted change in the treatment regime.

```
D[2] <- 1
(Y <- D * x * 1/sum(D * x))
```

```
 [1] 0.17523886 0.16200158 0.12979041 0.00000000 0.07424733 0.00000000
 [7] 0.04509712 0.19235156 0.00000000 0.22127314
```

Loess figures in ggplot2

We will learn more about nonparametric regression later in the course. Plotting the loess regression may suggest a model specification that may not be immediately apparent. Take, for example, the relationship between birthweight and maternal age. We can read in the data and take a random sample of 5,000, just to speed things up a bit.

```
library(ggplot2)
data <- read.csv("../../../resources/ps1.csv")
var.names <- c("dbrwt", "dmage")
idx <- sample.int(nrow(data), 5000)
sm.data <- data[idx, var.names]
```

The loess model is a specific variant of locally weighted scatterplot smoothing regression analysis. Note that this is very similar to the lowess model, which are separated only by the smoothing technique. When all observations are given equal weight over the domain, both loess and lowess are equivalent to simple linear regression. However, locally weighted regression analysis place higher weight on "nearby" observations. The specification of what constitutes "nearby" observations and the method of weighting distinguishes the various models. Mostly, the following just presents examples that are useful for learning `ggplot2`. Fig. 1 presents the loess curve on its own, and Fig. 2 presents the same curve with the scatterplot points.

```
(p <- ggplot(sm.data, aes(x=dmage, y=dbrwt)) + geom_smooth(method = "loess", size = 1.5))
```

The curve in Fig. 1 suggests that the relationship between maternal age and infant birthweight may be quadratic. Mothers that are relatively young and old may give birth to lighter babies, potentially indicating poorer infant health. Moreover, the confidence intervals are wider toward the tails of the age distribution, since there are fewer observations in these age ranges. To examine the curve in context of the actual observations, we can add the actual data as a scatterplot in Fig. 2.

```
p + geom_point()
```

The quadratic relationship is interesting, and we can show that it is statistically significant, but even for the small subsample of the data, the loess model explains only a small portion of the variation in birthweight.

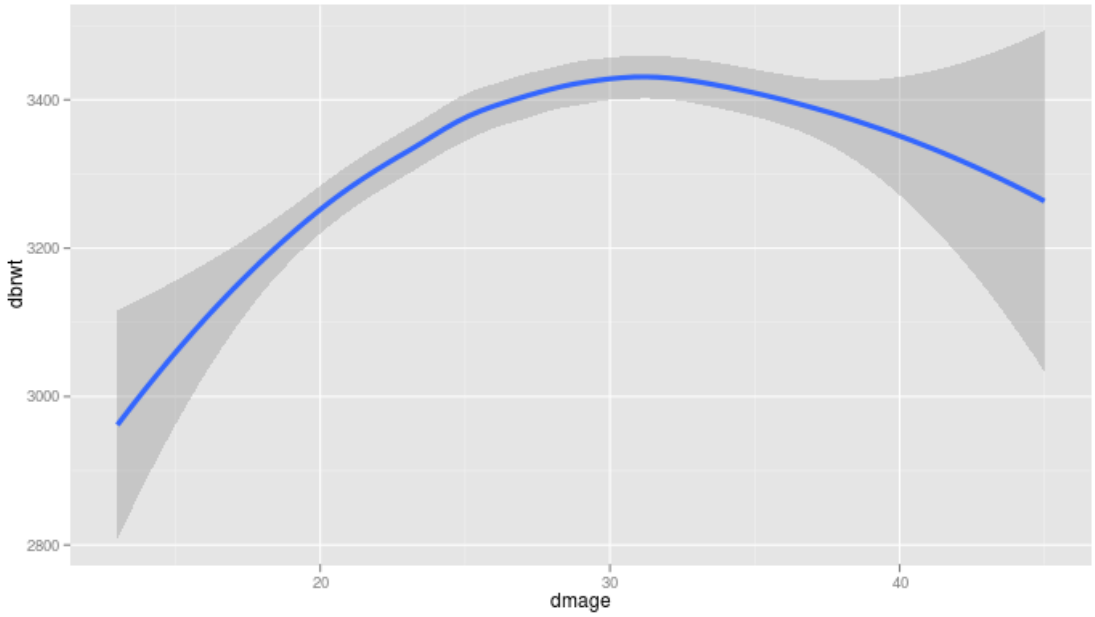

Figure 1: Loess regression

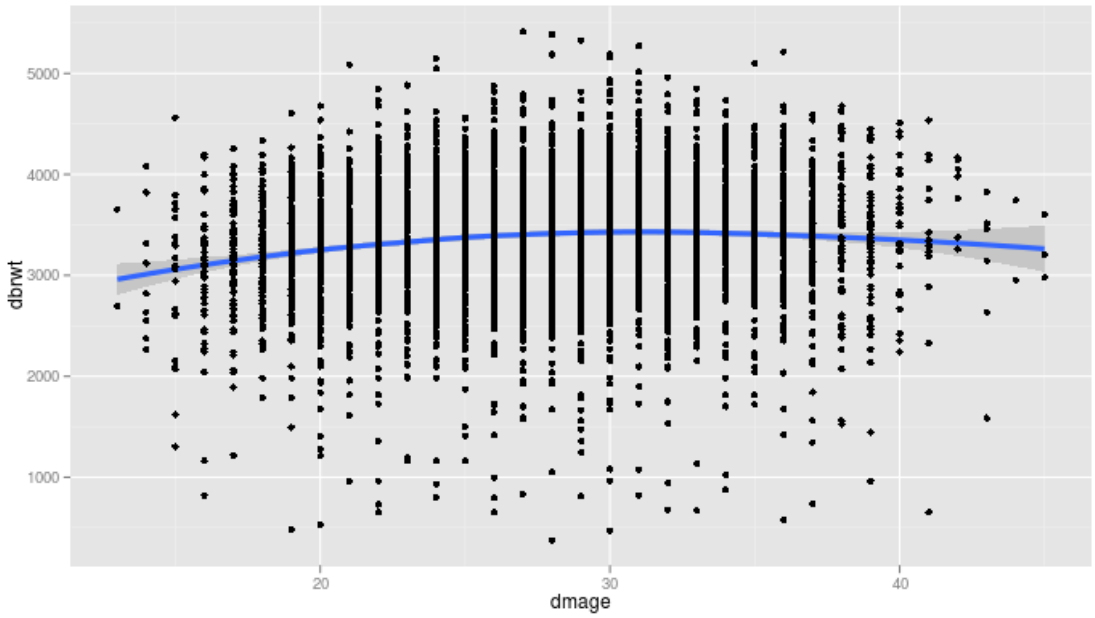

Figure 2: Loess regression with scatter plot