Load tidyverse libraries.

```
library(ggplot2)
library(tidyr)
library(dplyr)
```

Read in data, which is just the Excel sheet saved as a CSV.

```
data <- read.csv("Data624_project1_data.csv",header=TRUE)
```

Remove last rows where all fields are blank.

```
data <- data[1:9732,]
```

Convert to long format.

```
data_gathered <- gather(data,
    key="Variable",
    value="Value",
    -SeriesInd,-group)
data_gathered <- data.frame(data_gathered,
    Group.plus.var = paste0(data_gathered$group,"\n",data_gathered$Variable),
    stringsAsFactors=FALSE)
data_gathered$Group.plus.var <- factor(data_gathered$Group.plus.var,
    levels=paste0(rep(paste0("S0",1:6),each=5),"\n",rep(c("Var01","Var02","Var03","Var0
5","Var07"),times=6)))
```
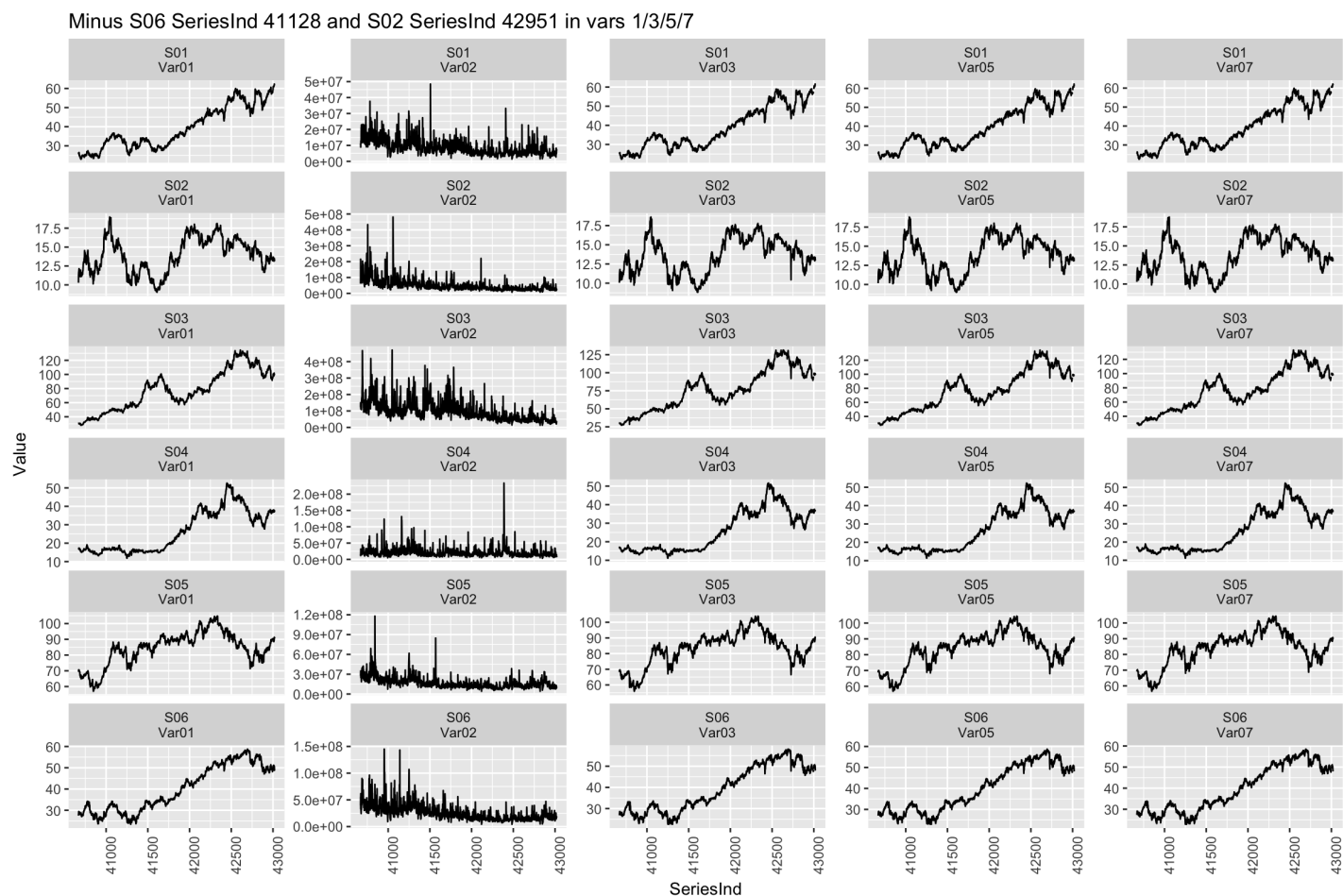
Make line plots.

```
ggplot(data_gathered,
    aes(SeriesInd,Value)) +
    geom_line() +
    facet_wrap(~Group.plus.var,scales="free_y",nrow=6,ncol=5) +
    theme(axis.text.x=element_text(angle=90, hjust=1))
```
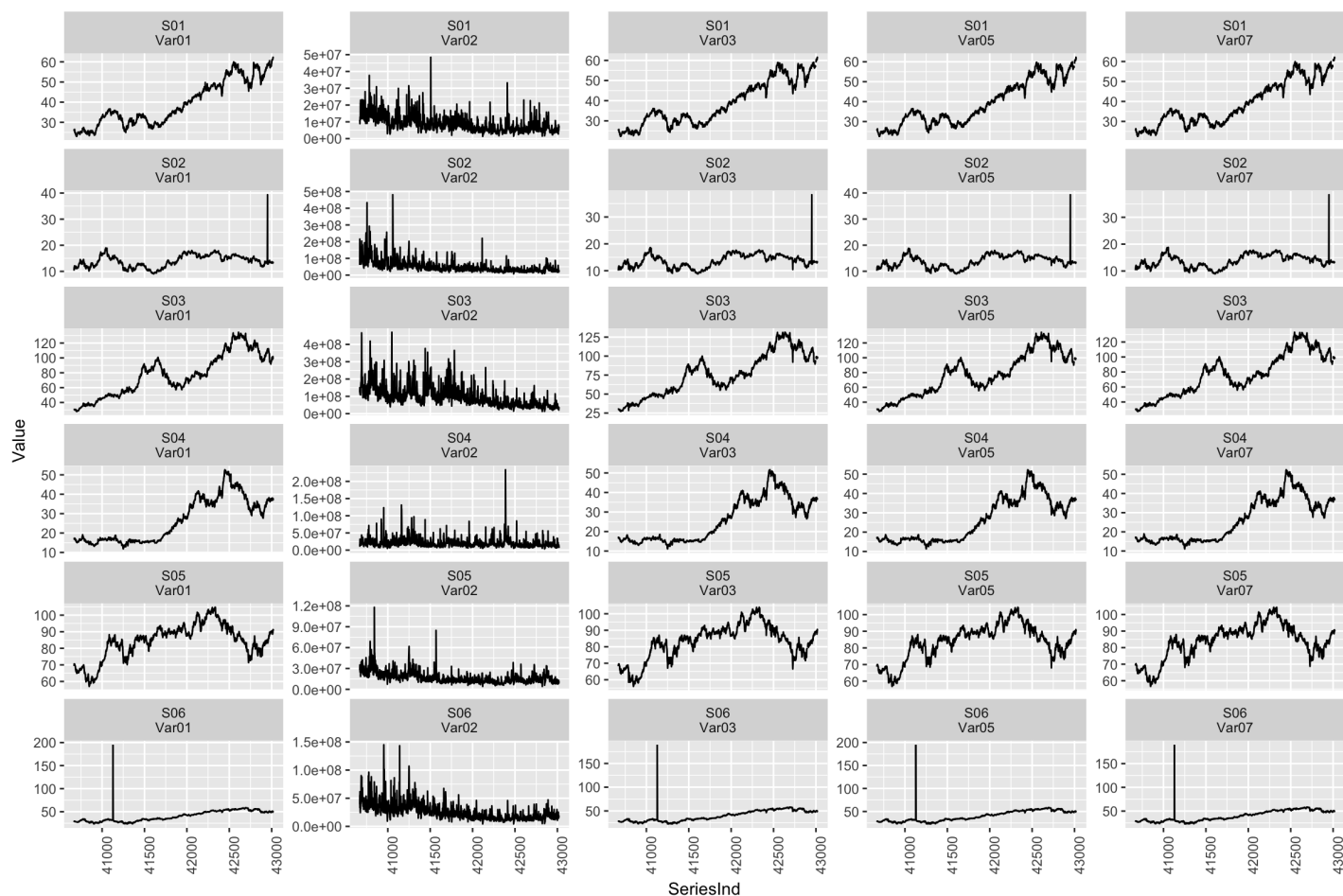
Plot minus S06 consistent outliers (SeriesInd 41128, variables 1,3, 5 and 7) and minus S02 consistent outliers (SeriesInd 42951, variables 1, 3, 5, and 7).

```
extreme_outliers <- which(data_gathered$SeriesInd == 41128 & data_gathered$group == "S0
6" & data_gathered$Variable != "Var02")
extreme_outliers <- c(extreme_outliers,
    which(data_gathered$SeriesInd == 42951 & data_gathered$group == "S02" & data_gathere
d$Variable != "Var02"))
data_gathered_minus_outliers <- data_gathered[setdiff(1:nrow(data_gathered),extreme_outl
iers),]
```

```
ggplot(data_gathered_minus_outliers,
    aes(SeriesInd,Value)) +
    geom_line() +
    facet_wrap(~Group.plus.var,scales="free_y",nrow=6,ncol=5) +
    theme(axis.text.x=element_text(angle=90, hjust=1)) +
    ggtitle("Minus S06 SeriesInd 41128 and S02 SeriesInd 42951 in vars 1/3/5/7")
```

Minus S06 SeriesInd 41128 and S02 SeriesInd 42951 in vars 1/3/5/7

We now see more clearly that variables 1, 3, 5, and 7 tend to have a similar time pattern within each series.

Let's try plotting these variables on the same plot for each series.

```
ggplot(data_gathered_minus_outliers[data_gathered_minus_outliers$Variable != "Var02",],
    aes(SeriesInd,Value,colour=Variable)) +
    geom_line() +
    facet_wrap(~group,scales="free_y",nrow=3,ncol=2) +
    theme(axis.text.x=element_text(angle=90, hjust=1)) +
    ggtitle("Minus S06 SeriesInd 41128 and S02 SeriesInd 42951 in vars 1/3/5/7")
```

Minus S06 SeriesInd 41128 and S02 SeriesInd 42951 in vars 1/3/5/7
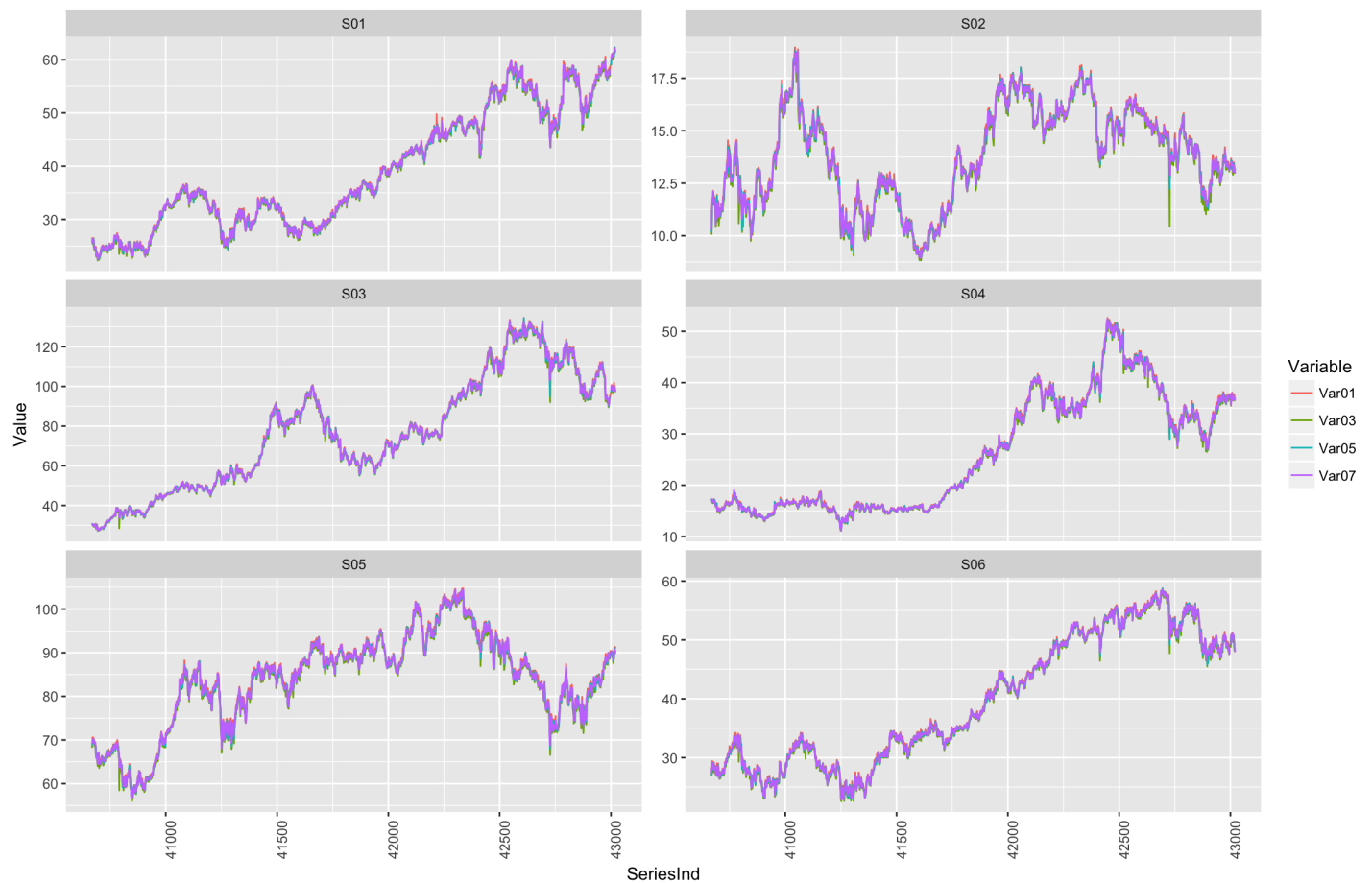
Now, plot Var02 alone.

```
ggplot(data_gathered[data_gathered$Variable == "Var02",],
    aes(SeriesInd,Value)) +
    geom_line() +
    facet_wrap(~group,scales="free_y",nrow=3,ncol=2) +
    theme(axis.text.x=element_text(angle=90, hjust=1)) +
    ggtitle("Var02")
```

Var02