DataVizA Tutorial: Additional Issues: Solutions

Department of Econometrics and Business Statistics, Monash University

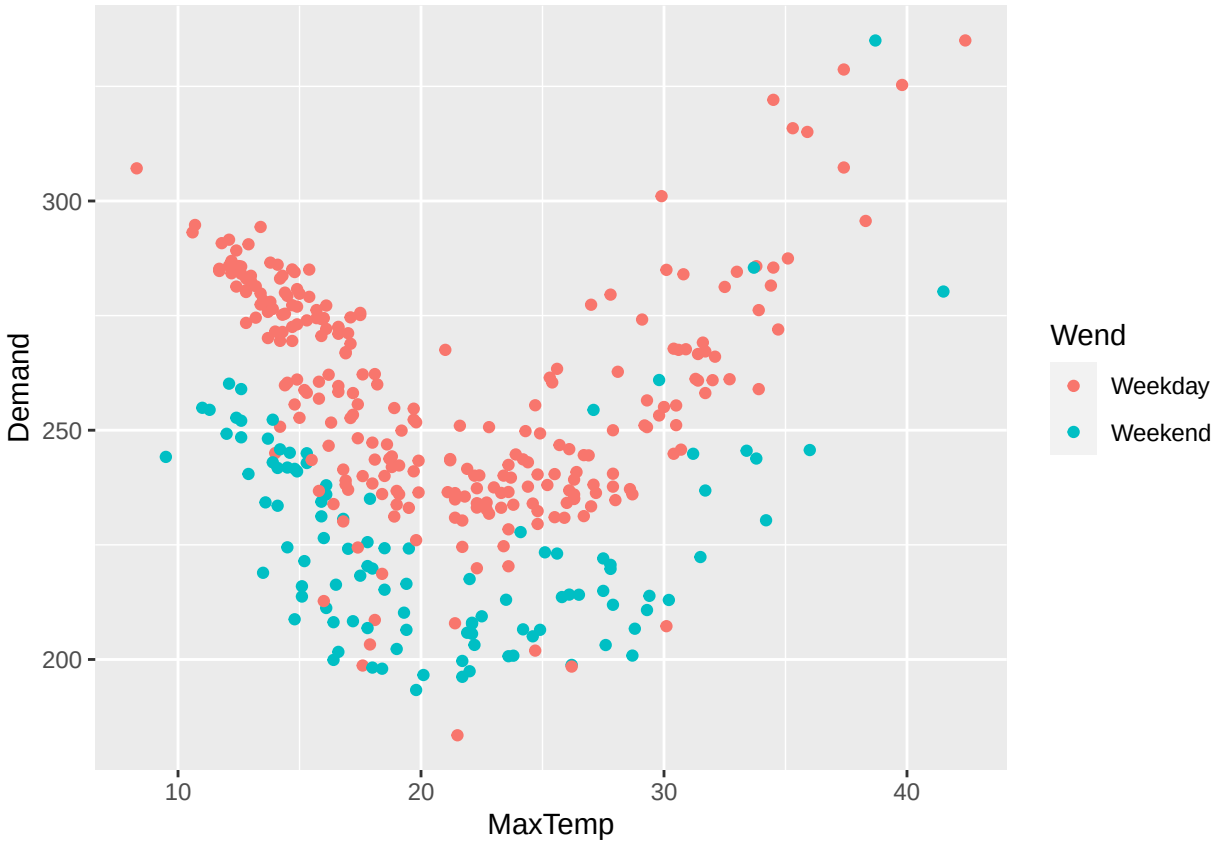
Tutorial 7

Electricity Data

The file *electricity.csv* contains data on the total daily demand of electricity in Victoria, the maximum daily temperature and an indicator variable for weekend/weekday.

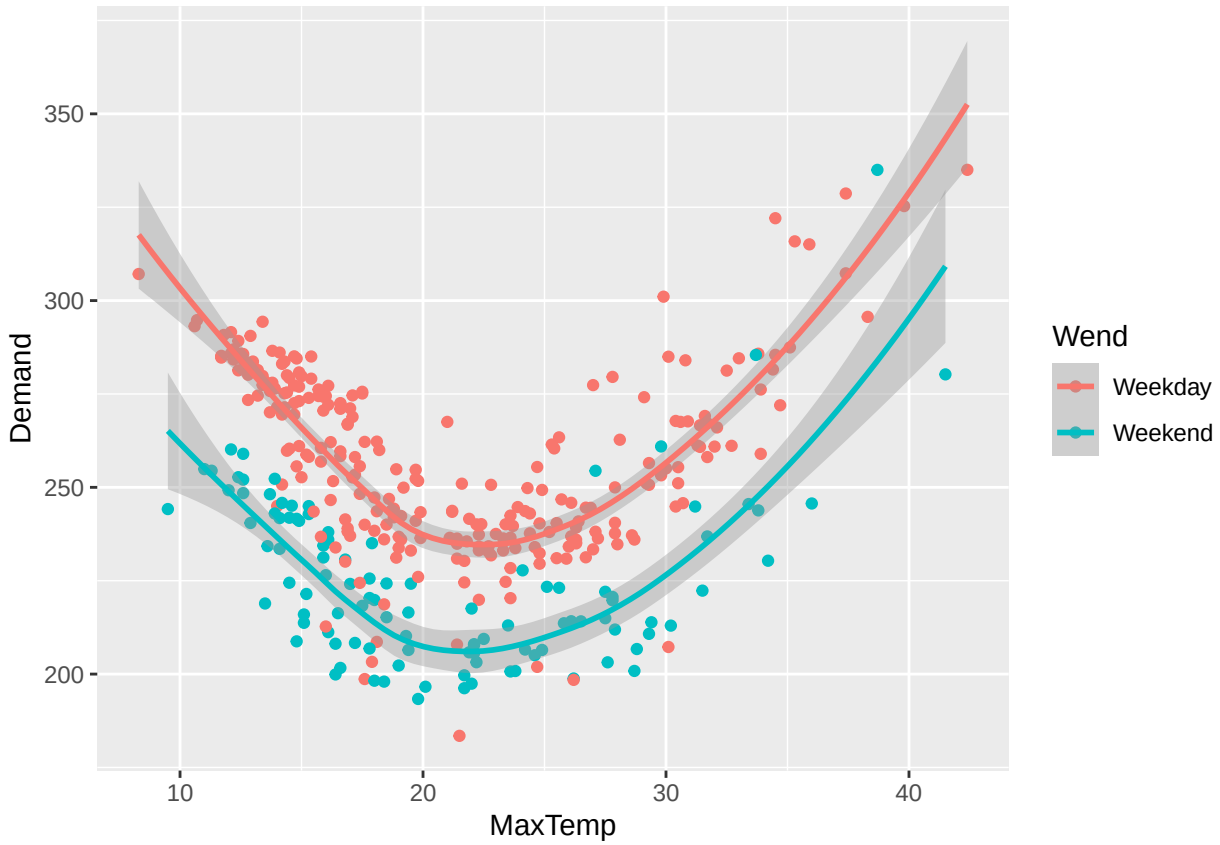
1. Plot a scatter plot of Maximum daily temperature (x axis) against total daily demand (y axis) with weekends and weekdays in different colors.

```
library(tidyverse)
elec<-read_csv('electricity.csv')

ggplot(elec,aes(x=MaxTemp,y=Demand,col=Wend))+geom_point()
```



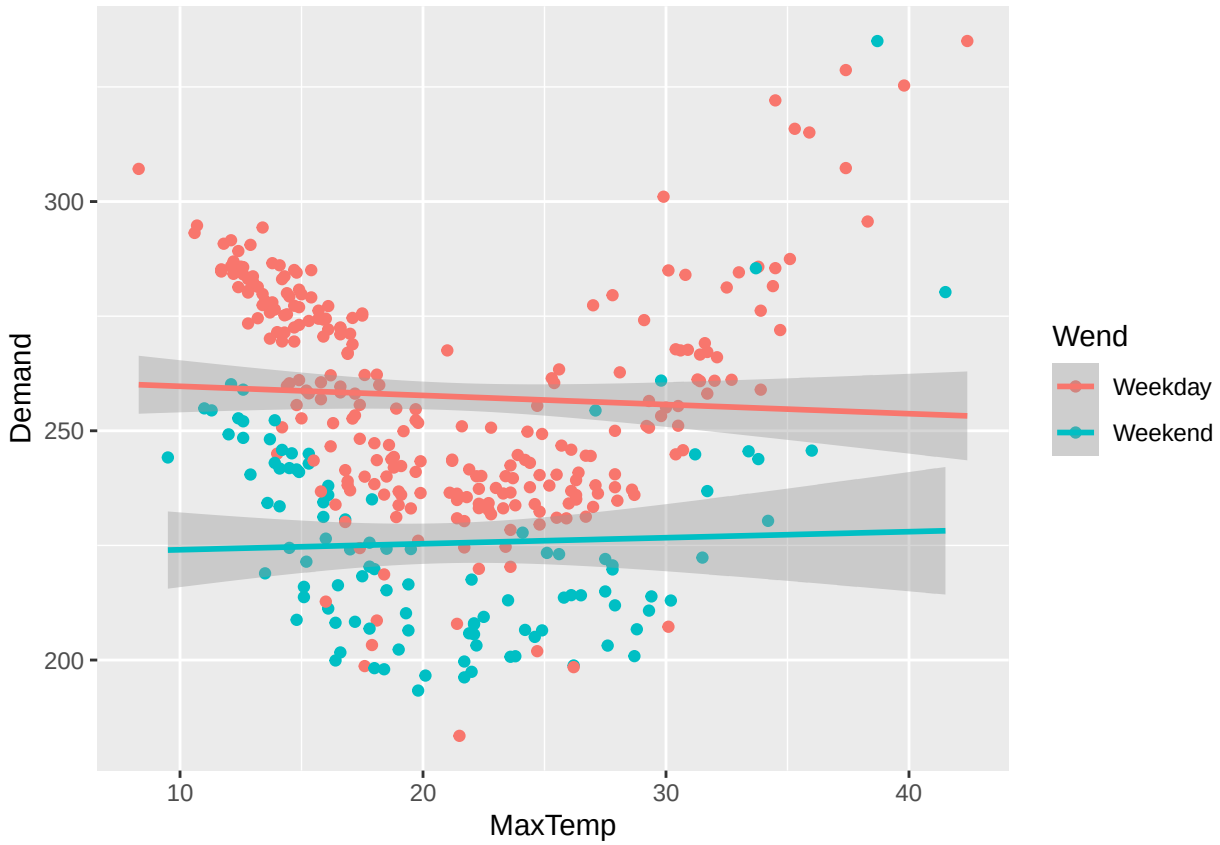
2. Add smooth lines obtained by a LOESS estimate.

```
ggplot(elec,aes(x=MaxTemp,y=Demand,col=Wend))+
  geom_point()+geom_smooth()
```



3. Add the line of best fit from a linear regression model.

```
ggplot(elec,aes(x=MaxTemp,y=Demand,col=Wend))+
  geom_point()+geom_smooth(method = 'lm')
```



4. What regression model could be suggested by such a plot?

Certainly not a linear mode. A model with a quadratic term would be a good starting point. Also it is interesting to note that the weekend seems to shift the intercept. Therefore a including a dummy variable for the weekend variable and no interaction terms should be a good baseline model.

5. Using plotly investigate the outlier that is a work day but has a very low consumption of electricity. Why may this be.

```
#The interactive plot will not render on a pdf but the following #code works.
library(plotly)
g<-ggplot(elec,aes(x=MaxTemp,
                   y=Demand,
                   col=Wend,
                   text=Date))+geom_point()
ggplotly(g)

#By hovering over the outlier you should see that the date is
#December 25th 2017.  Energy demand is lower on public holidays
#since most workplaces are closed and many people travel.
```

Swiss Exports (One last time)

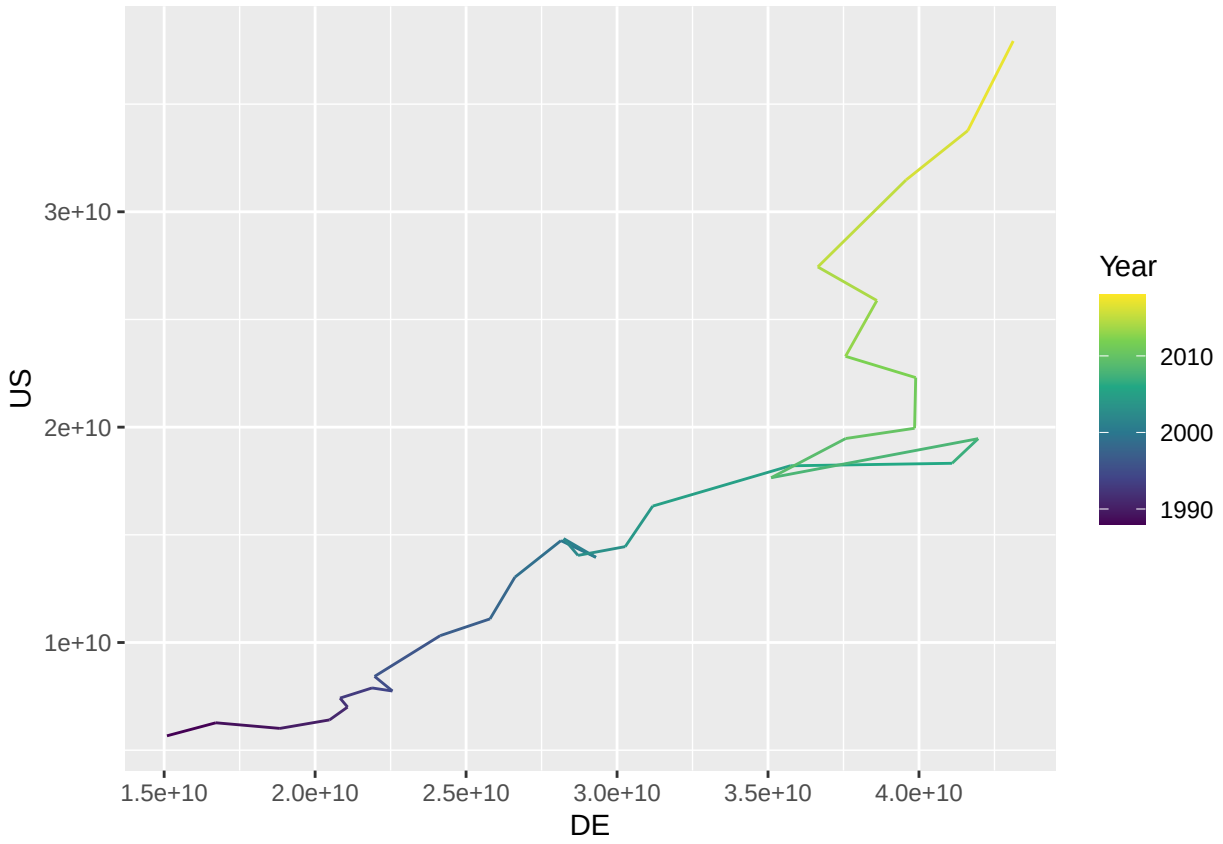
6. Create a path plot of the yearly exports to Germany against the yearly exports to the USA.

```
#Revision from last week

#Import data
```

```r
SwissWide<-read_csv('SwissExportsFull.csv')%>%
  rename(`NA`=X154)
```

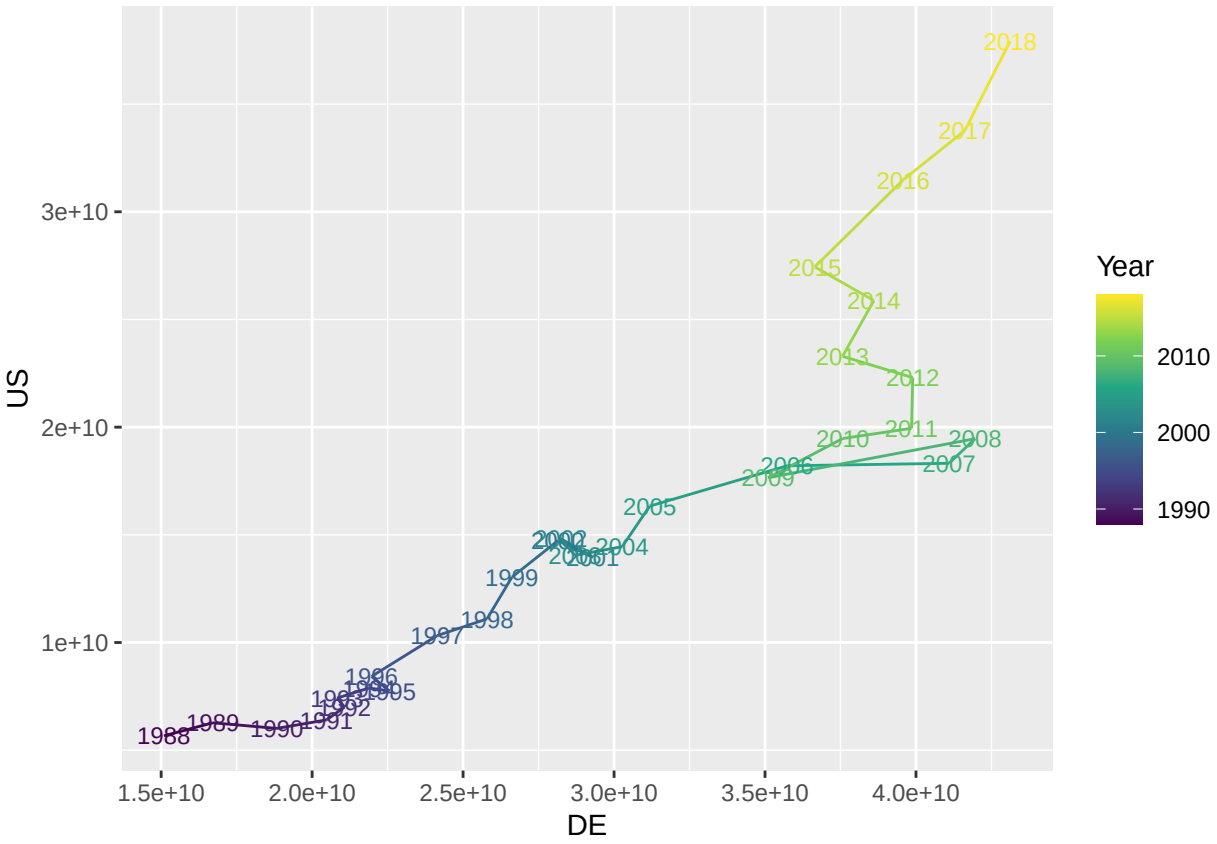
```
## Warning: Missing column names filled in: 'X154' [154]
```

```r
#Convert to long form
SwissLong<-pivot_longer(data = SwissWide,
                        cols = c(-Date, -Year), #Do not use these variables
                        names_to = 'Country', #Column names become variable
                        values_to = 'Exports') #All numbers are exports


#Create yearly aggregate
SwissLong%>% #This is much easier with long data
  select(-Date)%>% #Eventually cannot aggregate dates
  group_by(Year,Country)%>%
  summarise(YearlyExports=sum(Exports))->SwissYearly

#Select Germany and US
SwissYearly%>%
  filter(Country %in% c('DE','US'))%>%
  pivot_wider(id_cols = Year,
              names_from = Country,
              values_from = YearlyExports)->Exports_DE_US

#Finally make plot

ggplot(Exports_DE_US,aes(x=DE,y=US,col=Year))+
  geom_path()+scale_color_viridis_c()
```



7. Add the year labels as text to the plot

```
ggplot(Exports_DE_US,aes(x=DE,y=US,col=Year,label=Year))+
  geom_path()+geom_text(size=3)+scale_color_viridis_c()
```



8. What do you see in this plot?

Up until the GFC exports are trending upwards for both countries at similar rates. There is a dramatic reveral is exports to both countries in 2009. For a period between 2011 and 2015 the slope reverses direction signalling a shift from Germany to the US as an export location. This may be associated with depreciations of the Euro (relative to the Swiss Franc) that made Swiss goods too expensive for German consumers. Since 2015 however, exports to both the US and Germany has continued to grow.