

Analysis of Tidycorn

Rachel Feingold

2/27/2017

2. This data is considered a time series, want to analyze it as such.
3. Looked at the SALES from 5 States in different areas of the country

First, we need to load the data frame tidycorn, which was already tidied. We tidied the corn.csv file in tidycorn2.R

```
load("tidycorn.Rdata")
load("corndog.sav")
library(ggplot2)
library(lubridate)
```

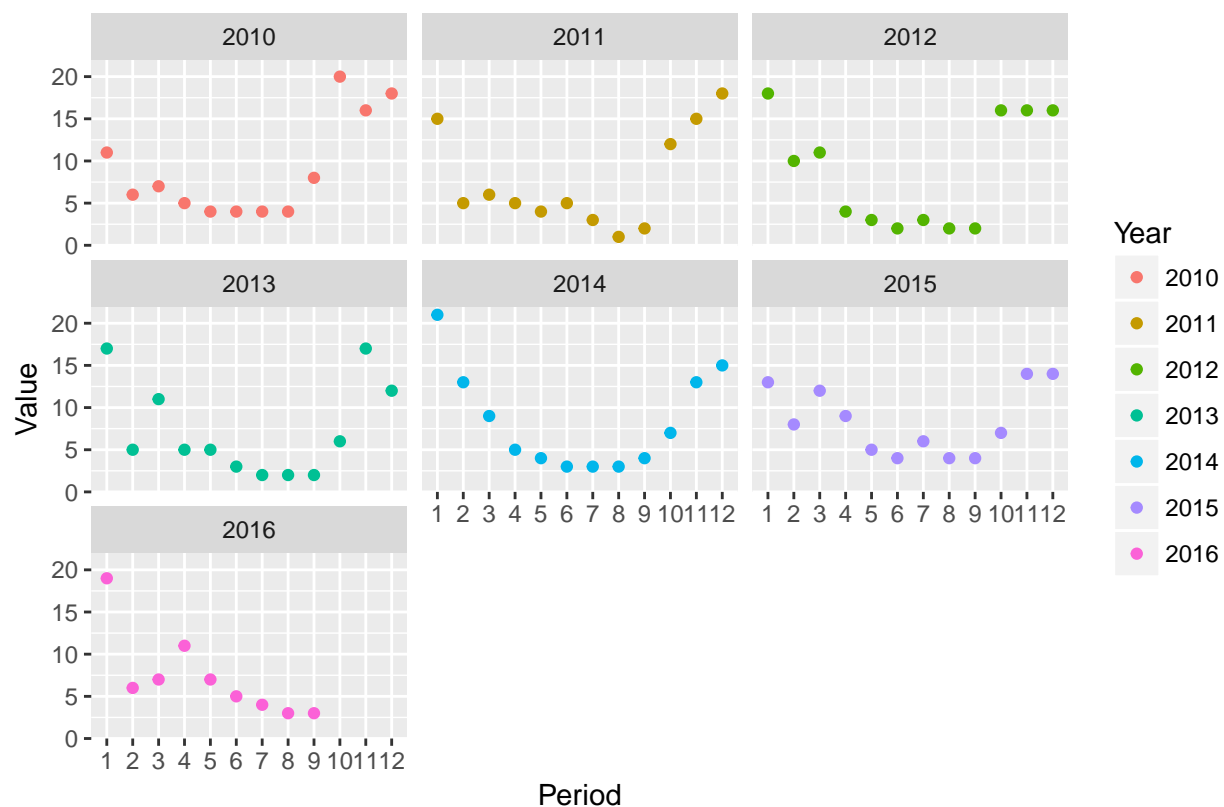
```
##
## Attaching package: 'lubridate'
## The following object is masked from 'package:base':
##
##      date
SALES = subset(tidycorn, subset = Data.Item.sub2=="SALES")
```

Tidycorn has 8 variables: Year, Period (Month), State, State.ANSI (assigned a number to each state), Data.Item.sub2, Data.Item3, Data.Item4, and Value

First, we'll look at SALES from Colorado. SALES is a factor level of Data.Item.sub2 and Colorado refers to the State variable. To begin, we need to take a subset, SALES of the data frame tidycorn. This subset includes all the observations where Data.Item.sub2=="SALES" Next, we take another subset, CO_SALES, which is a subset of all observations where the State is "colorado" Graphically, we made a plot of the monthly corn SALES in percentage of the market year, for the years 2010-2016. In other words, we are looking at which months of the year did we make the most sales. The second plot creates a regression line, for each year, of the monthly Corn SALES. This plot helps us better understand the long term, monthly trends of the corn sales over 7 years. The last plot looks at sales over time (ymd function transforms period) Consequently, I did this for 4 other states, in different regions of the United States.

```
CO_SALES = subset(SALES, subset = State=="colorado")
ggplot(CO_SALES) + geom_point(aes(x=Period, y=Value, col=Year)) + facet_wrap(~Year) + ggtitle("CORN - SAL")
```

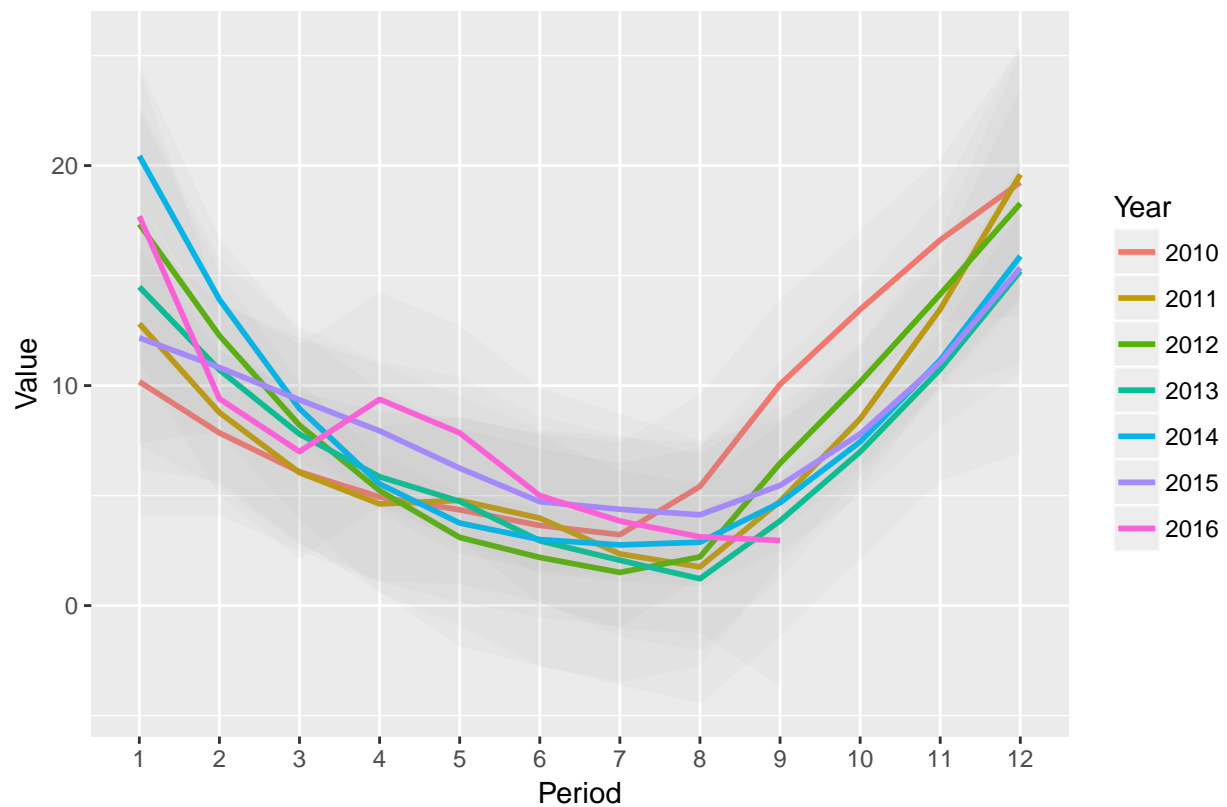
CORN – SALES IN PCT OF MKTG YEAR – COLORADO



```
ggplot(CO_SALES) + geom_smooth(aes(x=Period,y=Value,group=Year,col=Year),alpha = 0.05) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – COLORADO")
```

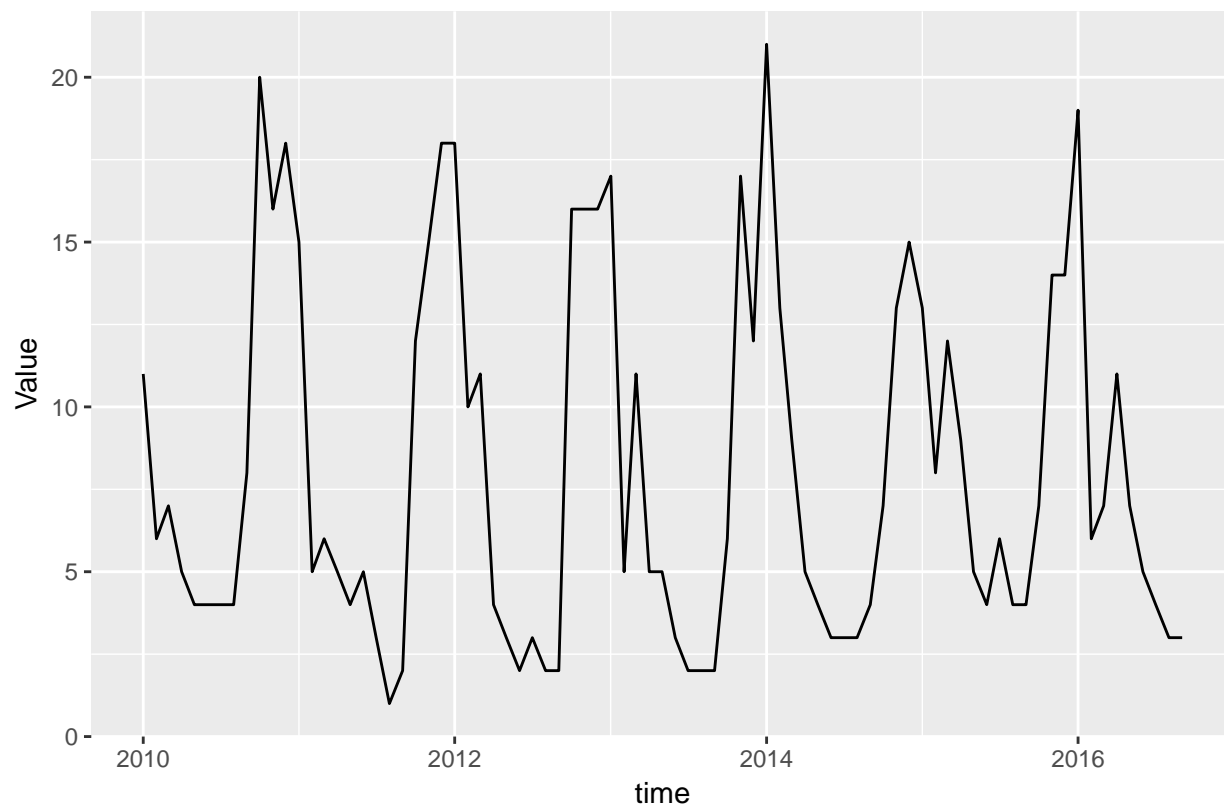
```
## `geom_smooth()` using method = 'loess'
```

CORN – SALES IN PCT OF MKTG YEAR – COLORADO



```
CO_SALES$time = paste(CO_SALES$Year,CO_SALES$Period,"01",sep = "-")
CO_SALES$time = ymd(CO_SALES$time)
ggplot(CO_SALES) + geom_line(aes(x = time,y = Value)) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – COLORADO")
```

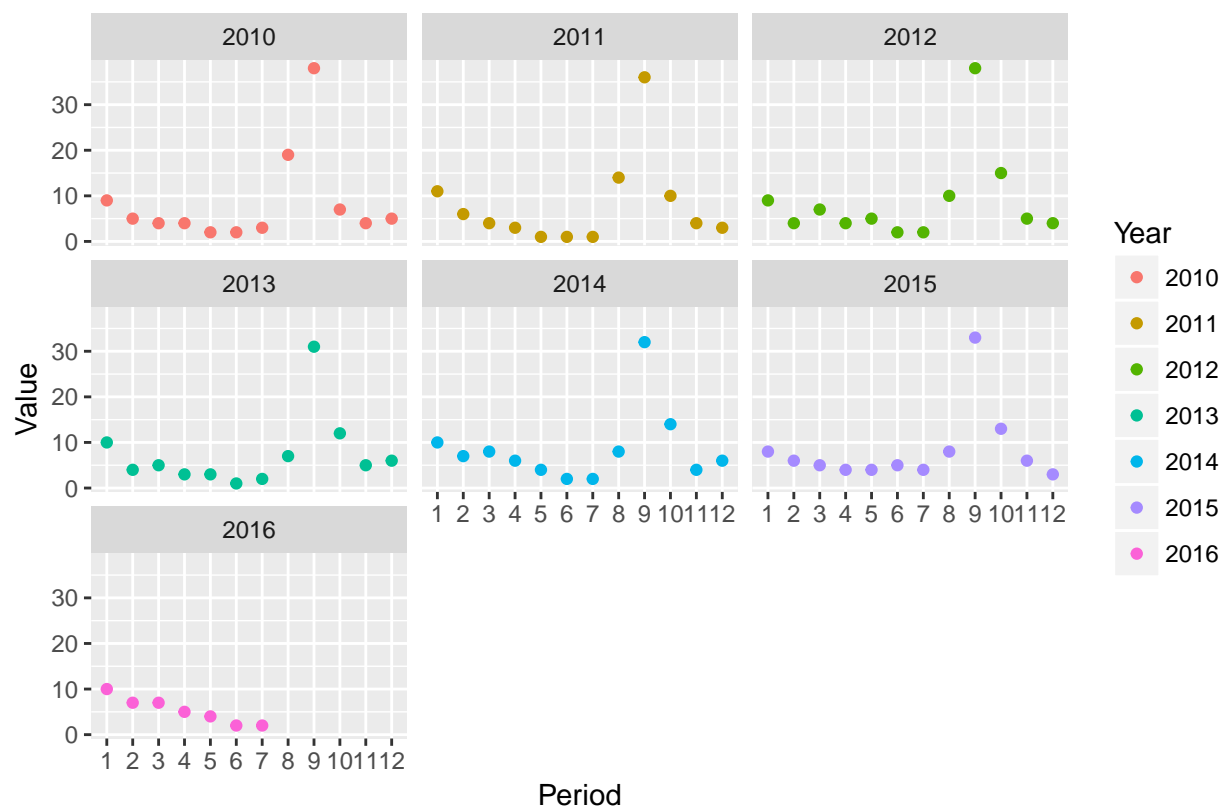
CORN – SALES IN PCT OF MKTG YEAR – COLORADO



North Carolina:

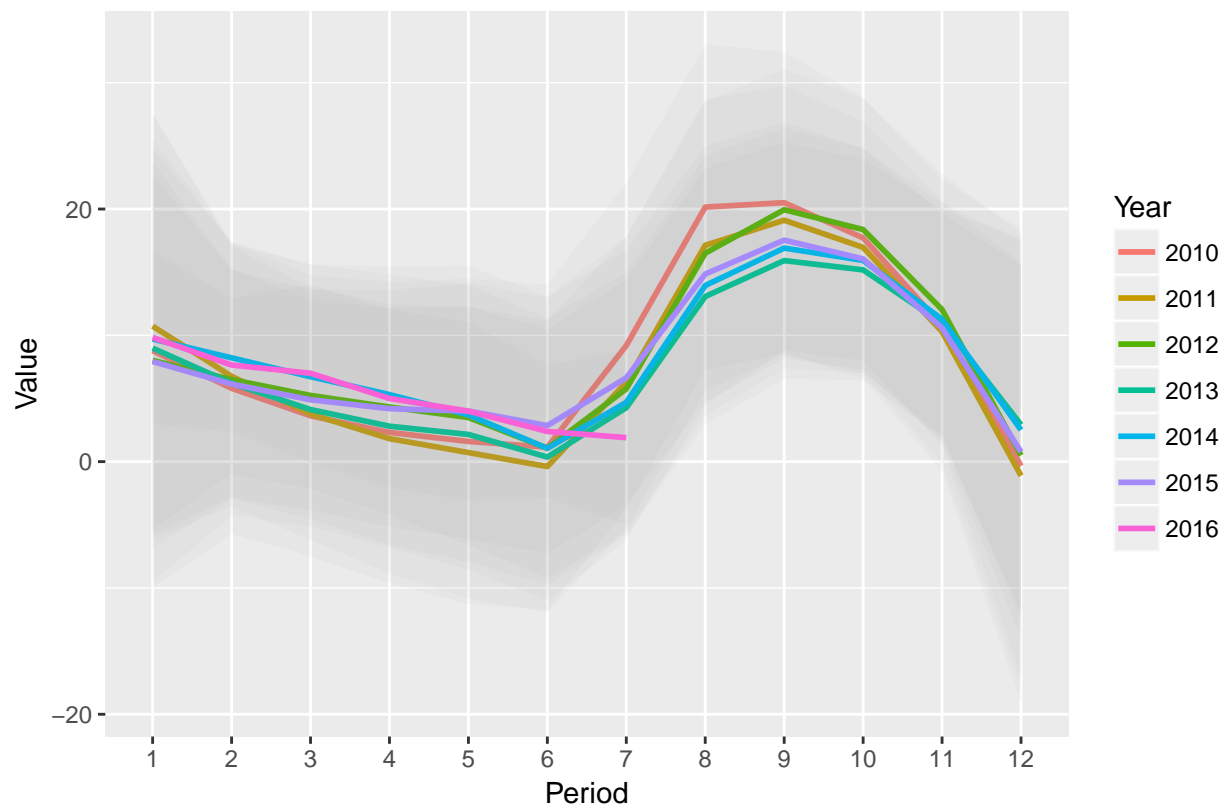
```
SALES = subset(tidycorn,subset = Data.Item.sub2=="SALES")
NC_SALES = subset(SALES,subset = State=="north carolina")
ggplot(NC_SALES) + geom_point(aes(x=Period,y=Value,col=Year)) + facet_wrap(~Year) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – NORTH CAROLINA")
```

CORN – SALES IN PCT OF MKTG YEAR – NORTH CAROLINA



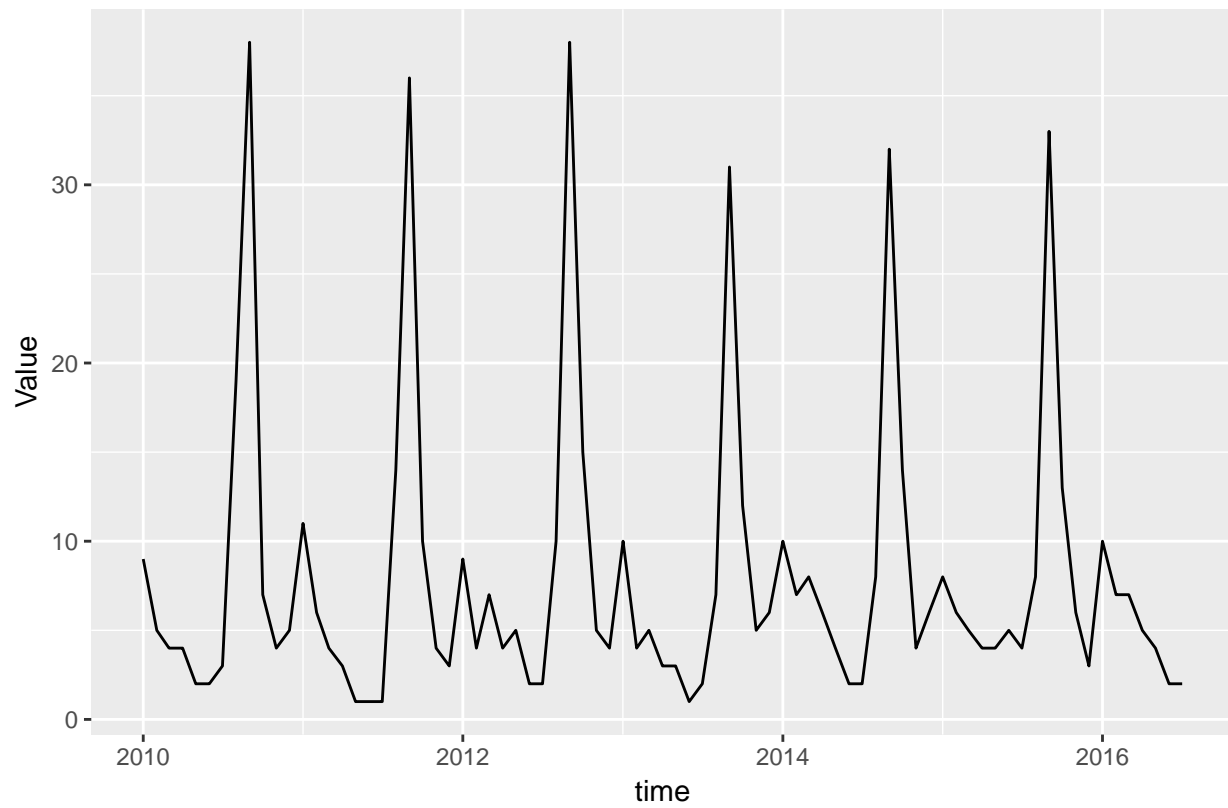
```
ggplot(NC_SALES) + geom_smooth(aes(x=Period,y=Value,group=Year,col=Year),alpha = 0.05) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – NORTH CAROLINA")
## `geom_smooth()` using method = 'loess'
```

CORN – SALES IN PCT OF MKTG YEAR – NORTH CAROLINA



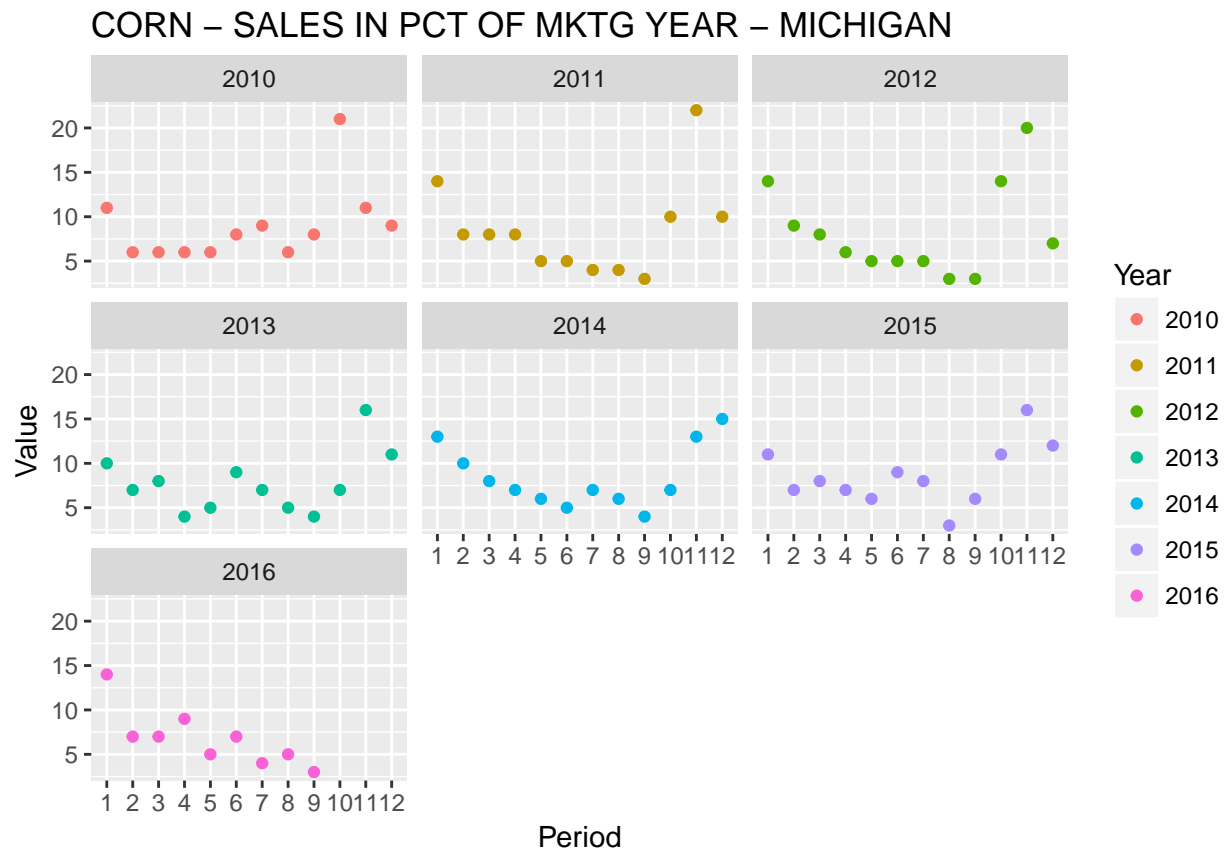
```
NC_SALES$time = paste(NC_SALES$Year,NC_SALES$Period,"01",sep = "-")
NC_SALES$time = ymd(NC_SALES$time)
ggplot(NC_SALES) + geom_line(aes(x = time,y = Value)) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – NORTH CAROLINA")
```

CORN – SALES IN PCT OF MKTG YEAR – NORTH CAROLINA

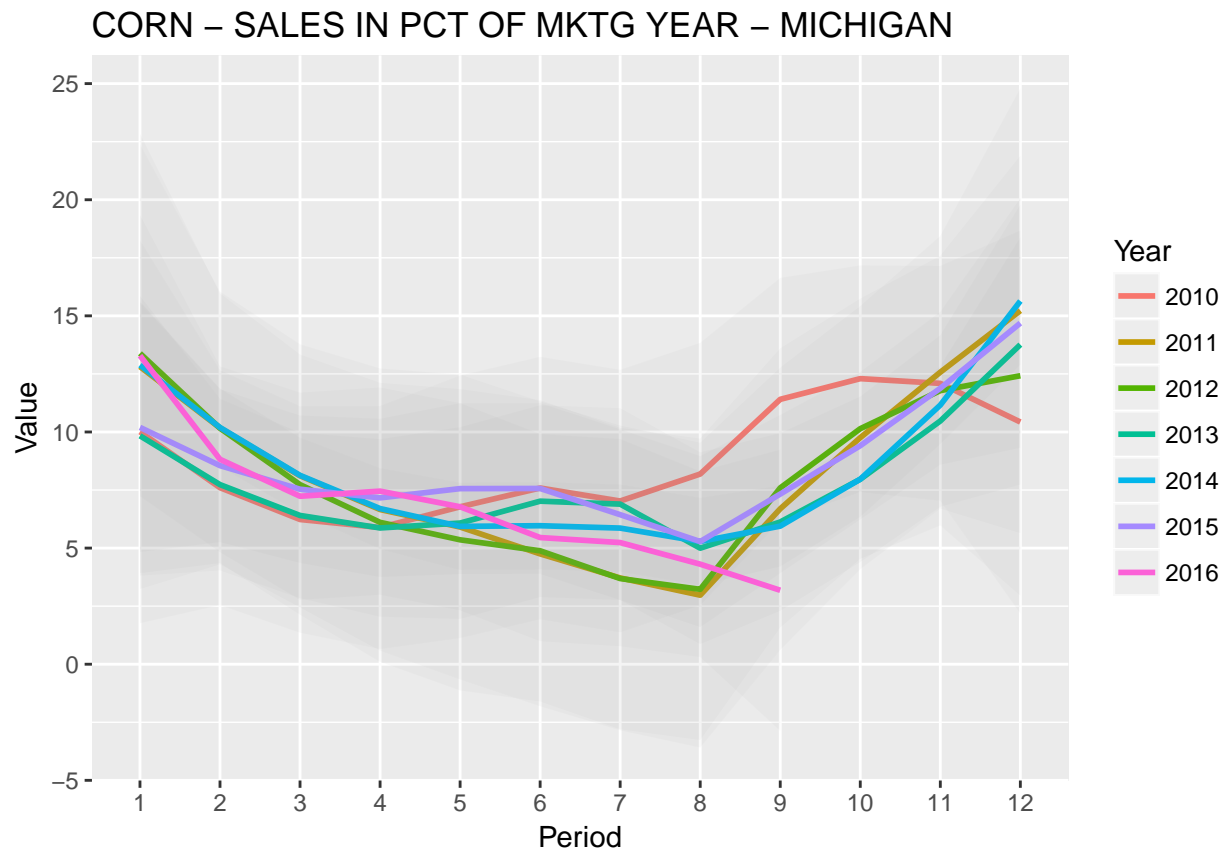


Michigan:

```
SALES = subset(tidycorn,subset = Data.Item.sub2=="SALES")
MI_SALES = subset(SALES,subset = State=="michigan")
ggplot(MI_SALES) + geom_point(aes(x=Period,y=Value,col=Year)) + facet_wrap(~Year) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – MICHIGAN")
```

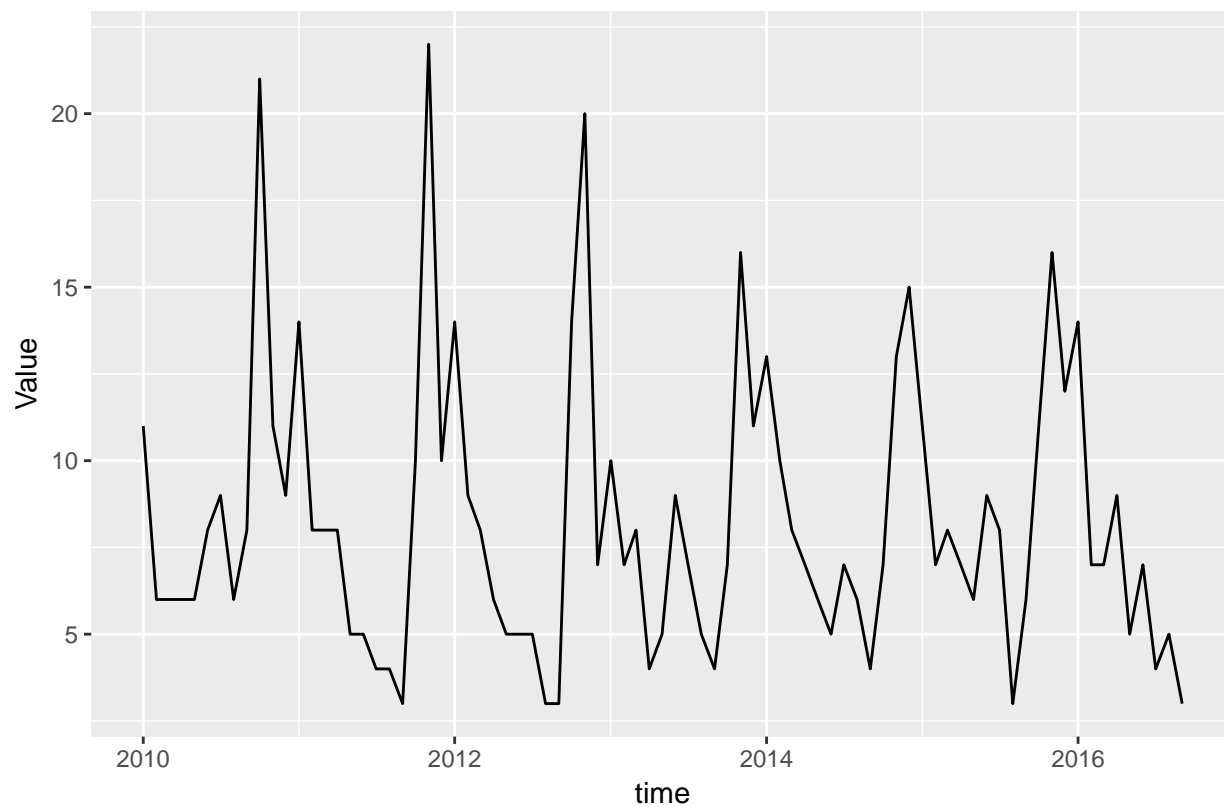


```
ggplot(MI_SALES) + geom_smooth(aes(x=Period,y=Value,group=Year,col=Year),alpha = 0.05) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – MICHIGAN")
## `geom_smooth()` using method = 'loess'
```

```
MI_SALES$time = paste(MI_SALES$Year,MI_SALES$Period,"01",sep = "-")
MI_SALES$time = ymd(MI_SALES$time)
ggplot(MI_SALES) + geom_line(aes(x = time,y = Value)) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – MICHIGAN")
```

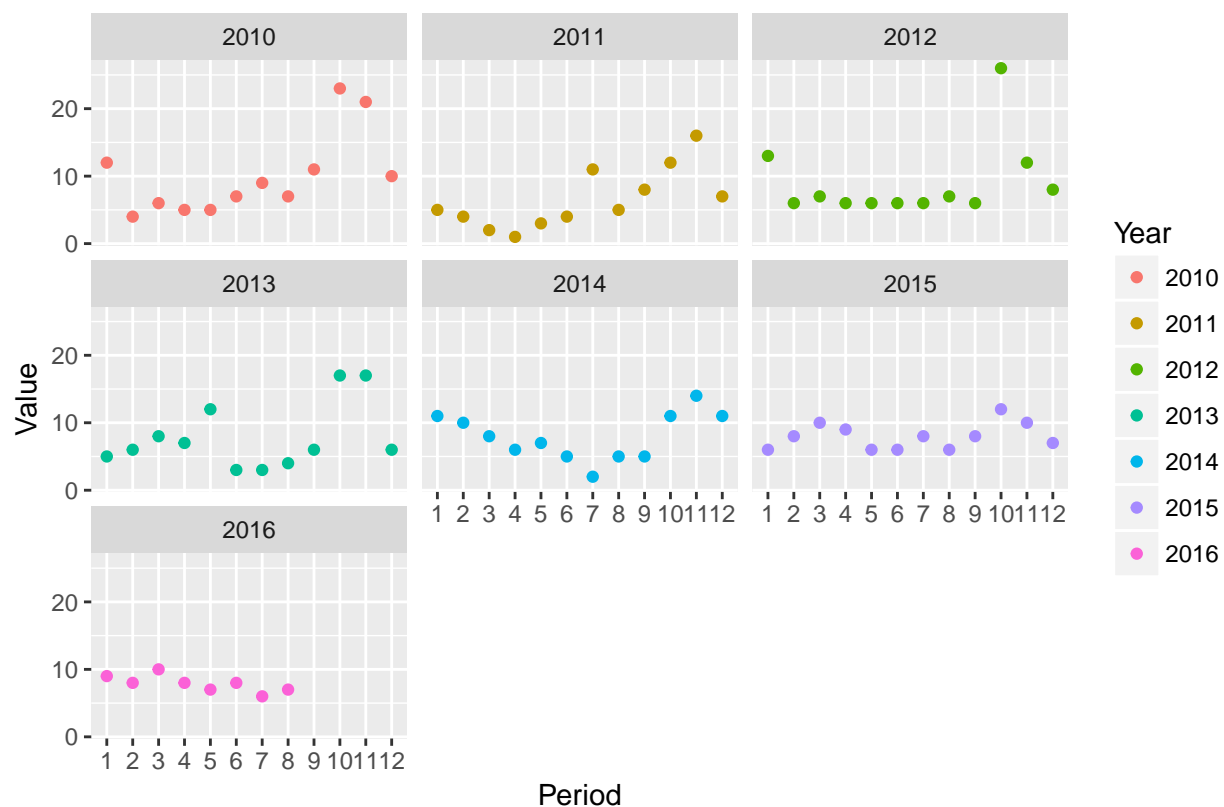
CORN – SALES IN PCT OF MKTG YEAR – MICHIGAN



Pennsylvania:

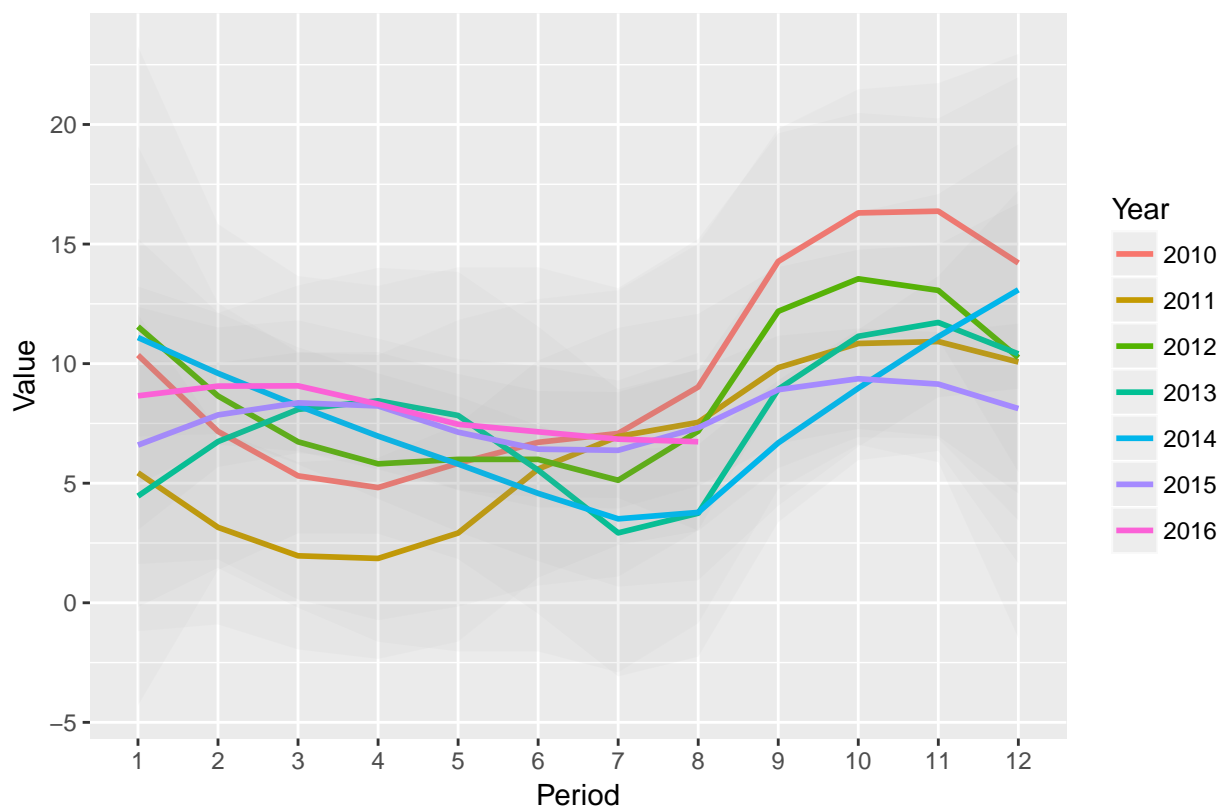
```
SALES = subset(tidycorn, subset = Data.Item.sub2=="SALES")
PA_SALES = subset(SALES, subset = State=="pennsylvania")
ggplot(PA_SALES) + geom_point(aes(x=Period, y=Value, col=Year)) + facet_wrap(~Year) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – PENNSYLVANIA")
```

CORN – SALES IN PCT OF MKTG YEAR – PENNSYLVANIA



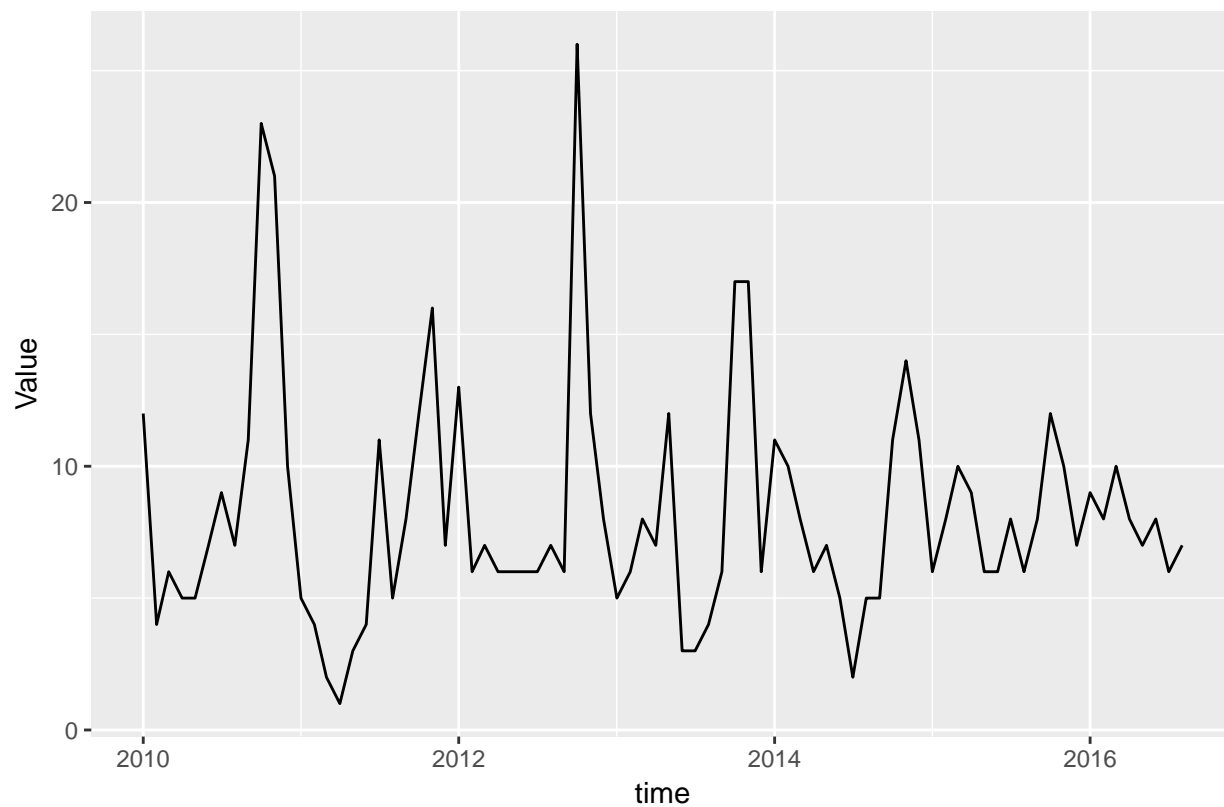
```
ggplot(PA_SALES) + geom_smooth(aes(x=Period,y=Value,group=Year,col=Year),alpha = 0.05) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – PENNSYLVANIA")
## `geom_smooth()` using method = 'loess'
```

CORN – SALES IN PCT OF MKTG YEAR – PENNSYLVANIA



```
PA_SALES$time = paste(PA_SALES$Year,PA_SALES$Period,"01",sep = "-")
PA_SALES$time = ymd(PA_SALES$time)
ggplot(PA_SALES) + geom_line(aes(x = time,y = Value)) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – PENN")
```

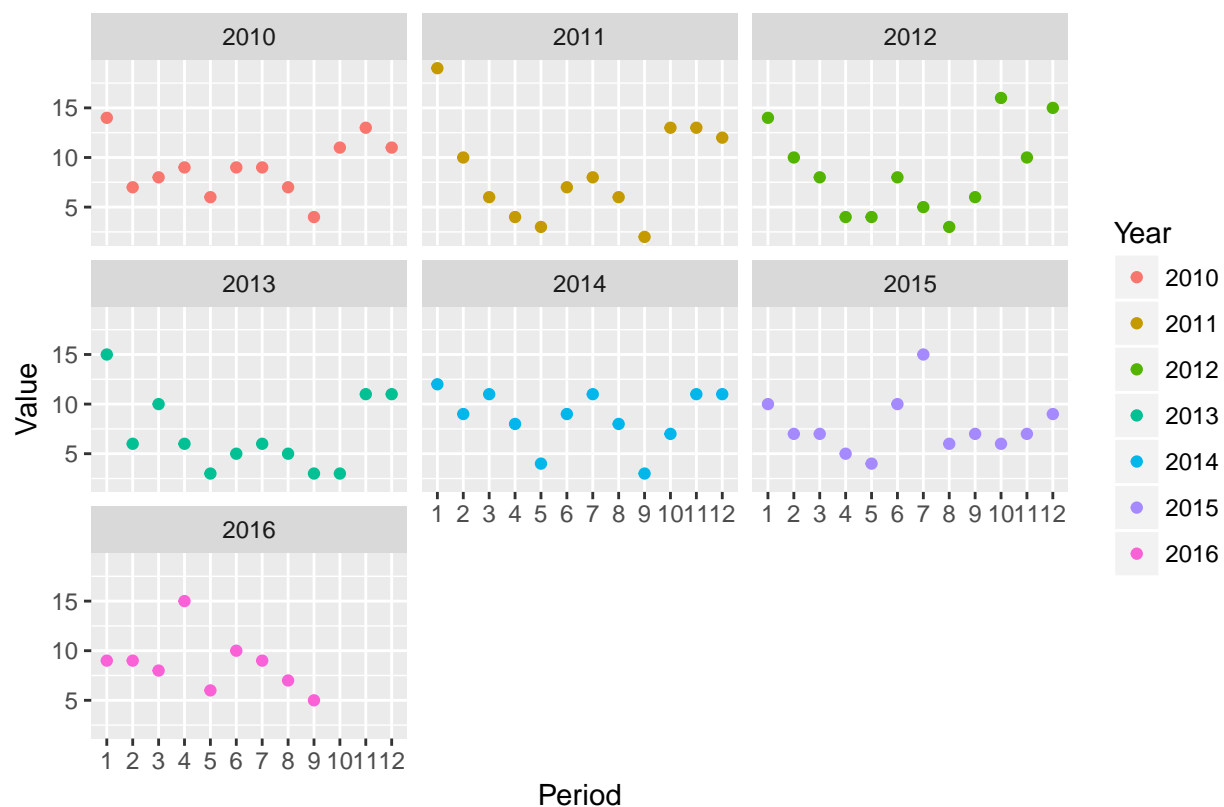
CORN – SALES IN PCT OF MKTG YEAR – PENNSYLVANIA



North Dakota

```
SALES = subset(tidycorn,subset = Data.Item.sub2=="SALES")
ND_SALES = subset(SALES,subset = State=="north dakota")
ggplot(ND_SALES) + geom_point(aes(x=Period,y=Value,col=Year)) + facet_wrap(~Year) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – NORTH DAKOTA")
```

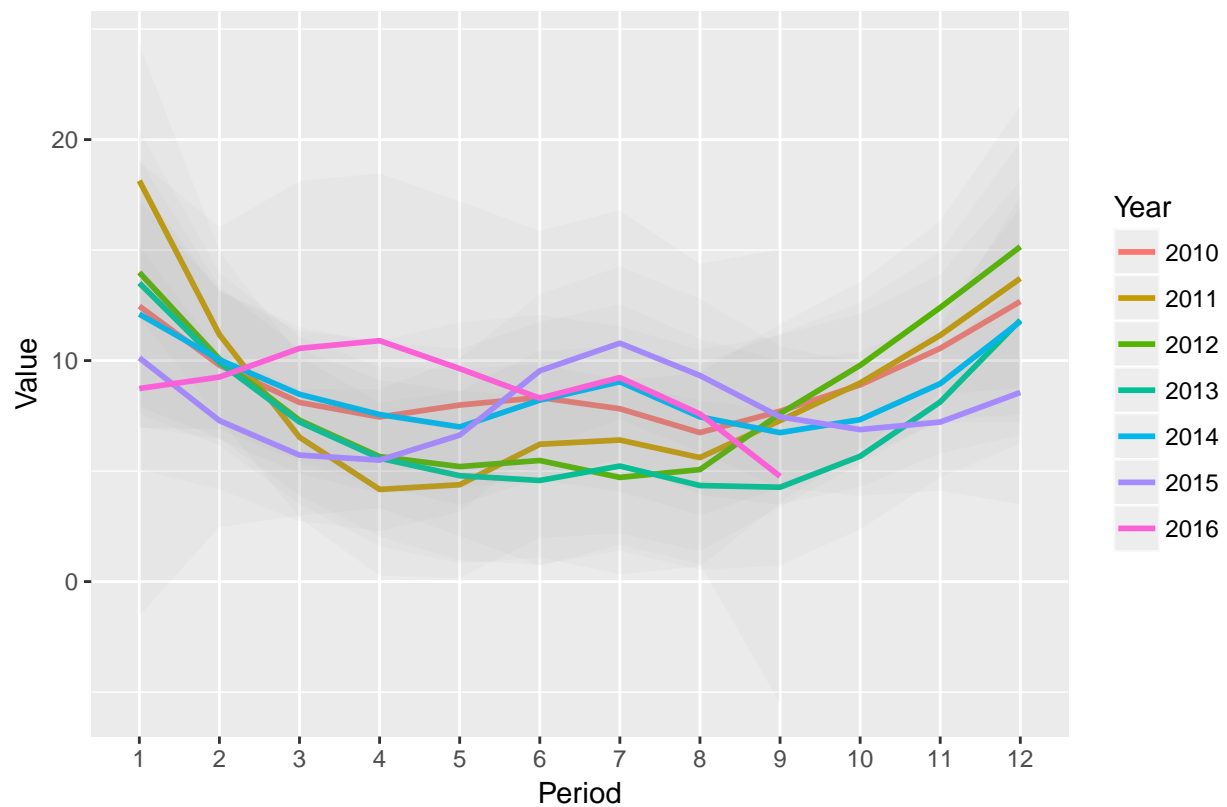
CORN – SALES IN PCT OF MKTG YEAR – NORTH DAKOTA



```
ggplot(ND_SALES) + geom_smooth(aes(x=Period,y=Value,group=Year,col=Year),alpha = 0.05) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – NORTH DAKOTA")
```

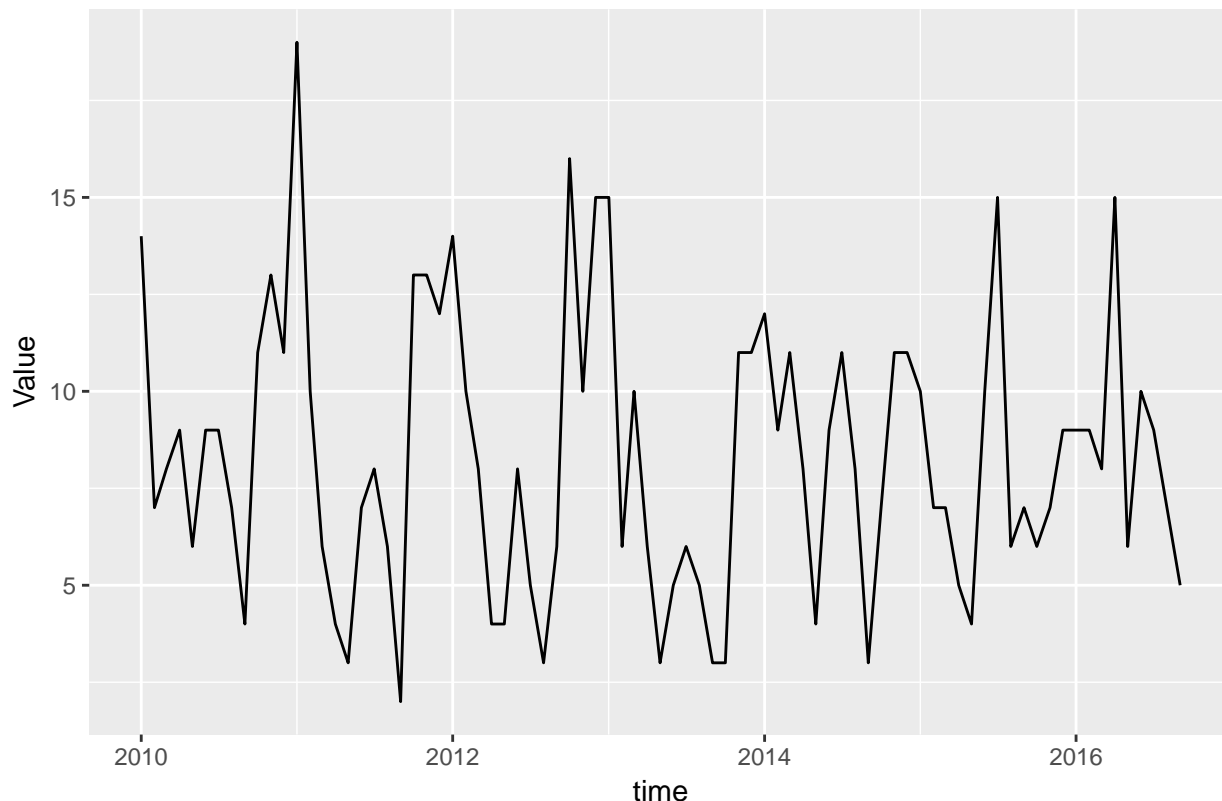
```
## `geom_smooth()` using method = 'loess'
```

CORN – SALES IN PCT OF MKTG YEAR – NORTH DAKOTA



```
ND_SALES$time = paste(ND_SALES$Year,ND_SALES$Period,"01",sep = "-")
ND_SALES$time = ymd(ND_SALES$time)
ggplot(ND_SALES) + geom_line(aes(x = time,y = Value)) + ggtitle("CORN – SALES IN PCT OF MKTG YEAR – NORTH DAKOTA")
```

CORN – SALES IN PCT OF MKTG YEAR – NORTH DAKOTA



Observations and Conclusions: In Colorado, looking at the monthly corn sales for each year plot, there is definitely a trend. The peak sales are from December to January and there is a crash in sales from June to August. The line of best fit is a parabola for all the years. When correlating all the years on a single plot, this trend is clearer. Examining the third plot, which shows the monthly corn sales over the years as if it is a continuous curve, even though these data points are discrete. For the most part, the corn sales are seasonal; the regression line is sinusoidal between time period and sales. In other words, this trend in sales increasing in the winter and decreasing in the summer is cyclical and consistent.

Conversely, in North Carolina, the significant peak month for corn sales is in September and there are rather stable sales for the rest of the year. The second plot mirrors this prediction and the third plot shows that the trend in sales is even more consistent through the years in North Carolina than in Colorado. This is probably a result of the weather and farming climate in the southern United States.

For Michigan, the plots and trends are similar to Colorado. For Pennsylvania and North Dakota, the plots and trends are rather similar to North Carolina.

Finally, the map plot shows which regions of the US have the most sales in corn (the Midwest) and which has the least (the mid Atlantic region).

```
{r}
```

```
library(maps) library(data.table) library(lubridate) library(mapproj) SALES %<>% as.data.table() SALES
= SALES[,list(Value=median(Value)),by=State] SALES #dplyr::group_by() #aggregate() #do the same
things #data.table do it fast
```

```
statmap = map_data("state") corn_map = merge(SALES,statmap,by.x = "State", by.y = "region",all.y =
T) corn_map = arrange(corn_map,order,group)
```

```
ggplot(corn_map,aes(x=long, y=lat, group=group)) + geom_polygon(linetype = 2, size = 0.1, colour
= "lightgrey",aes(fill = Value)) + expand_limits(x = corn_map$long,y = corn_map$lat) + coord_map(
"polyconic")
```



```
rm(statmap,t,theme_clean)
```