Beer Ingredients Plot

Tracy

January 4, 2017

Executive Summary

This analysis is to do a replicative study of shrin's playground on world's beer ingredients and alcohol and tabacco data from tax and trade bureau(TTB) (https://www.ttb.gov/).

The data comes in by month, year and following ingredients:

- Malt and malt products
- Corn and corn products
- Rice and rice products
- Barley and barley products
- Wheat and wheat products
- Sugar and syrups
- Dry hops
- Hops extracts
- Other ingredients, like flavors, etc

Read in data

First of all, let's read in the data and order it by year and calculate the percentage change from previous month.

```
library(ttbbeer)
data("beermaterials")

library(dplyr)
# str(beermaterials)
# order by year from least to the most recent
beermaterials_change=arrange(beermaterials,Year)
# value=beermaterials_change[[colnames(beermaterials_change)[3]]] #that's how you get your value. by in
# value[1]
# beermaterials_change[,3]#this will show the whole list value of that column
for (i in 3:ncol(beermaterials_change)) {
  value=beermaterials_change[[colnames(beermaterials_change)[i]]]
  beermaterials_change[,i]=c(0,100*round((value[-1]-value[1:(length(value)-1)])/value[1:(length(value)-

  #the above formula (value[-1]-value[1:(length(value)-1)])/value[1:(length(value)-1)] is trying to get
  #this is to substitute the column value for the percentage
}

# head(beermaterials_change)

beermaterials$sum=rowSums(beermaterials[,-c(1:2)])
```

Second, let's change the data format from wide to long, in other words, simplify the data layout by just including 'ingredients' and 'value'.

```r
library(tidyr)
# gather is great to change wide format to long format.
beermaterials_gather=beermaterials %>%
  gather(Ingredients,Amount, Malt_and_malt_products:Other)
beermaterials_gather$Date=paste('01',beermaterials_gather$Month, beermaterials_gather$Year,sep = "-") %
  as.Date(format="%d-%B-%Y")
beermaterials_gather$Ingredients=gsub("_"," ",beermaterials_gather$Ingredients)

beermaterials_change_gather=beermaterials_change %>%
  gather(Ingredients,Amount, Malt_and_malt_products:Other)
beermaterials_change_gather$Date=paste('01',beermaterials_change_gather$Month, beermaterials_change_gat
  as.Date(format="%d-%B-%Y")
beermaterials_change_gather$Ingredients=gsub("_"," ",beermaterials_change_gather$Ingredients)
```

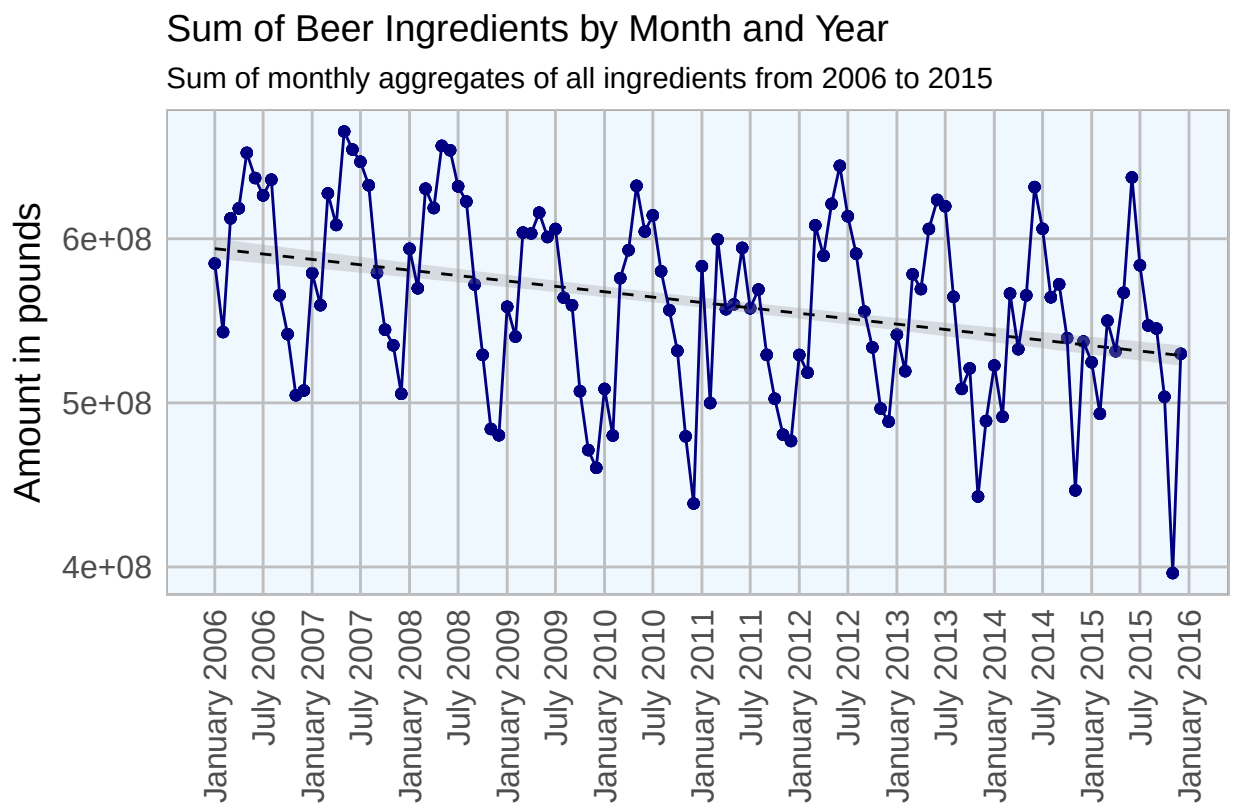
Exploratory analysis

Now let's do some exploratory analysis to see what the data looks like. Here we also set up the theme for the entire analysis.

First, we'd like to see the general trend of the beer ingredients by lumping them all together by 6 month interval.

```r
# install.packages('ggplot2')
library(ggplot2)

my_theme=function(base_size=12,base_family="sans"){
  theme_minimal(base_size=base_size,base_family=base_family)+
    theme(
      axis.text=element_text(size=12),
      axis.text.x=element_text(angle=90,vjust=0.5,hjust=1),
      axis.title=element_text(size=14),
      panel.grid.major=element_line(color='grey'),
      panel.grid.minor=element_blank(),
      panel.background=element_rect(fill="aliceblue"),
      strip.background=element_rect(fill="royalblue",color='grey',size=1),
      strip.text=element_text(face='bold',size=12,color='white'),
      legend.position='bottom',
      legend.justification='top',
      legend.box='horizontal',
      legend.box.background=element_rect(colour='grey50'),
      legend.background=element_blank(),
      panel.border=element_rect(color='grey',fill=NA,size=0.5)

  )

}

theme_set(my_theme())
# install.packages('lattice')
library(lattice)
ggplot(beermaterials_gather,aes(x=Date,y=sum))+
  geom_point(size=1.5,alpha=0.6,color='navy')+
  geom_smooth(alpha=0.3,color='black',size=0.5,linetype='dashed',method = 'gam')+
```

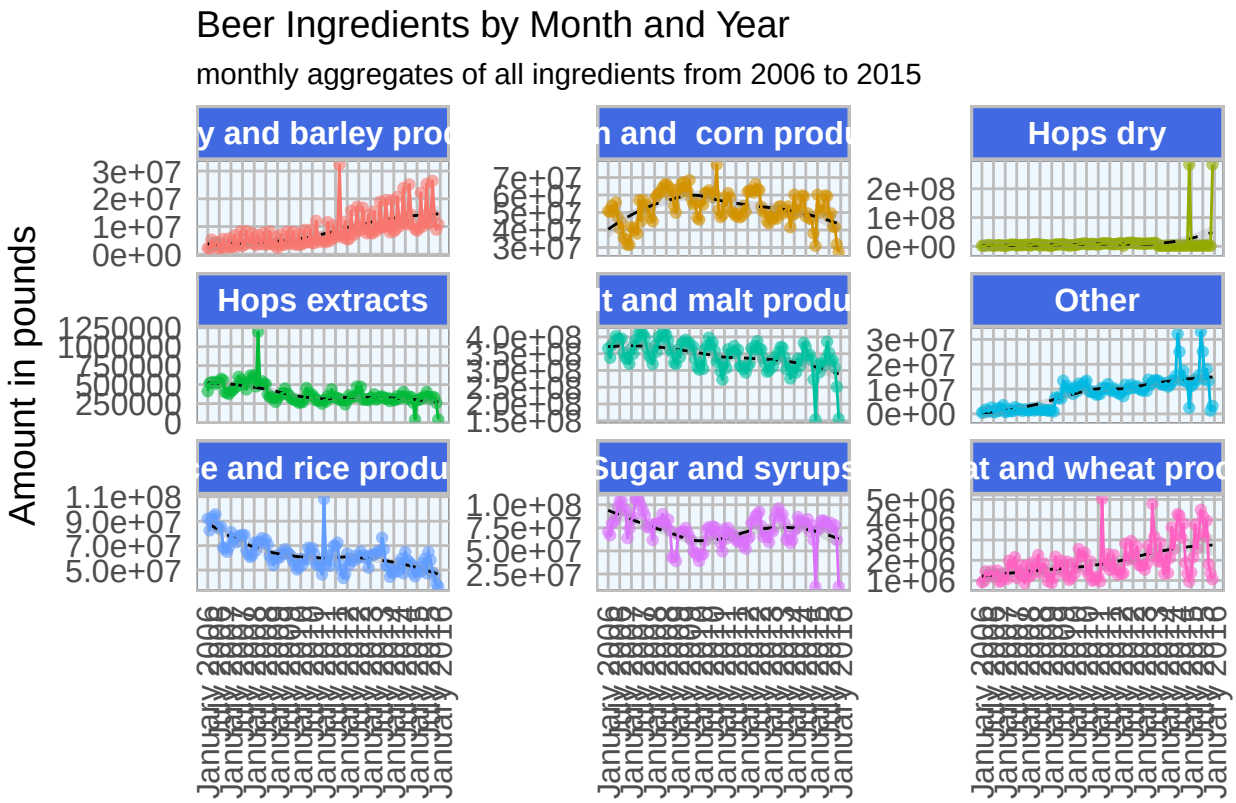
```
  geom_line(color='navy')+
  guides(color=FALSE)+
  labs(
    x="",
    y='Amount in pounds',
    title='Sum of Beer Ingredients by Month and Year',
    subtitle="Sum of monthly aggregates of all ingredients from 2006 to 2015"
  )+
  scale_x_date(date_breaks = '6 month',date_labels = "%B %Y")
```



Second, let's look at the amount of beer ingredients used over the time by 6 month interval:

```
ggplot(beermaterials_gather,aes(x=Date,y=Amount,colour=Ingredients))+
  geom_point(size=1.5,alpha=0.6)+
  geom_smooth(alpha=0.3,color='black',size=0.5,linetype='dashed')+
  geom_line()+
  guides(color=FALSE)+
  facet_wrap(~Ingredients,ncol = 3,scales = 'free_y')+
  labs(
    x="",
    y='Amount in pounds',
    title='Beer Ingredients by Month and Year',
    subtitle="monthly aggregates of all ingredients from 2006 to 2015"
  )+
  scale_x_date(date_breaks = '6 month',date_labels = "%B %Y") # date_breaks can specify any date interv
```

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



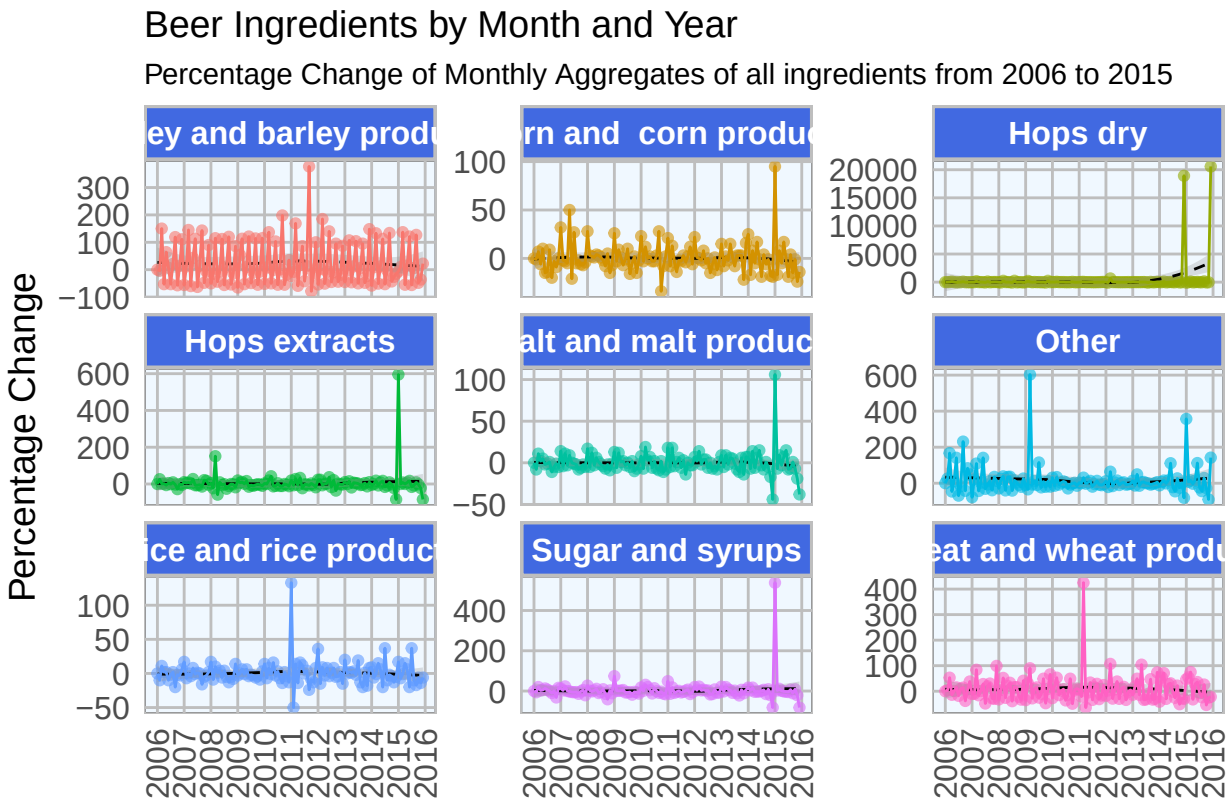
thirdly, let's zoom into more granular data by looking at it by 1 month interval and just display all the ingredients in one column:

```
ggplot(beermaterials_gather,aes(x=Date,y=Amount,colour=Ingredients))+
  geom_point(size=1.5,alpha=0.6)+
  # geom_smooth(alpha=0.3,color='black',size=0.5,linetype='dashed')+
  geom_line()+
  guides(color=FALSE)+
  facet_wrap(~Ingredients,ncol = 1,scales = 'free_y')+
  labs(
    x="",
    y='Amount in pounds',
    title='Beer Ingredients by Month and Year',
    subtitle="monthly aggregates of all ingredients from 2006 to 2015"
  )+
  scale_x_date(date_breaks = '1 month',date_labels = "%B %Y") # date_breaks can specify any date interv
```

fourthly, let's look at the monthly percentage change by different ingredients:

```
ggplot(beermaterials_change_gather,aes(x=Date,y=Amount,colour=Ingredients))+
  geom_point(size=1.5,alpha=0.6)+
  geom_smooth(alpha=0.3,color='black',size=0.5,linetype='dashed')+
  geom_line()+
  guides(color=FALSE)+
  facet_wrap(~Ingredients,ncol = 3,scales = 'free_y')+
  labs(
    x="",
    y='Percentage Change',
    title='Beer Ingredients by Month and Year',
    subtitle="Percentage Change of Monthly Aggregates of all ingredients from 2006 to 2015"
  )+
  scale_x_date(date_breaks = '1 year',date_labels = "%Y")
```

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



next, let's look at the percent change in pie format.

```
beeraterials_percent=cbind(beermaterials[,1:2],prop.table(as.matrix(beermaterials[,-c(1:2)]),margin=1)*
beermaterials_percent_gather=beeraterials_percent %>%
  gather(Ingredients,Amount, Malt_and_malt_products:Other)
beermaterials_percent_gather$Ingredients=gsub("_"," ",beermaterials_percent_gather$Ingredients)

#below is to make a customized factor for a certain variable
f=unique(beermaterials_percent_gather$Month)
beermaterials_percent_gather=within(beermaterials_percent_gather,{
  Month=factor(Month,levels = f)
  }
  )
```

Let's look at some animated charts:

```
library(animation)
ani.options(ani.width=800,ani.height=500)
saveGIF({
  for (year in rev(unique(beermaterials_percent_gather$Year))) {
    pie=ggplot(subset(beermaterials_percent_gather,Year==paste(year)),aes(x="",y=Amount,fill=Ingredients
      geom_bar(width=1,stat="identity")+theme_minimal()+coord_polar("y",start = 0)+
      labs(
        title="Percentage of Beer Ingredients by Year and Month",
        subtitle=paste(year))+
      #theme has to contain the below things but you can use element_blank() to mark empty
      theme(
```

```
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        panel.border = element_blank(),
        panel.grid = element_blank(),
        axis.ticks = element_blank(),
        plot.title = element_text(size = 14,face='bold'),
        legend.title = element_blank(),
        legend.position = "bottom",
        legend.text = element_text(size = 8))+
      facet_wrap(~Month,ncol = 6)+guides(fill=guide_legend(ncol = 9,byrow = F))
      print(pie)
  }
},movie.name = 'beer_ingredients.gif')
```

```
## Warning: running command 'C:\Windows\system32\cmd.exe /c convert --version'
## had status 4
```

```
## I cannot find ImageMagick with convert = "convert"
```

```
## Warning in normalizePath(path.expand(path), winslash, mustWork):
## path[1]="C:\Program Files\ImageMagick-7.0.4-Q16/convert.exe": The system
## cannot find the file specified
```

```
## but I can find it from the Registry Hive: C:\Program Files\ImageMagick-7.0.4-Q16
```

```
## Executing:
## ""C:\Program Files\ImageMagick-7.0.4-Q16\convert.exe" -loop 0
##     -delay 100 Rplot1.png Rplot2.png Rplot3.png Rplot4.png
##     Rplot5.png Rplot6.png Rplot7.png Rplot8.png Rplot9.png
##     Rplot10.png "beer_ingredients.gif""
```

```
## Warning: running command 'C:\Windows\system32\cmd.exe /c ""C:\Program
## Files\ImageMagick-7.0.4-Q16\convert.exe" -loop 0 -delay 100 Rplot1.png
## Rplot2.png Rplot3.png Rplot4.png Rplot5.png Rplot6.png Rplot7.png
## Rplot8.png Rplot9.png Rplot10.png "beer_ingredients.gif""' had status 1
```

```
## Warning in cmd.fun(convert): '""C:\Program Files\ImageMagick-7.0.4-
## Q16\convert.exe" -loop 0 -delay 100 Rplot1.png Rplot2.png Rplot3.png
## Rplot4.png Rplot5.png Rplot6.png Rplot7.png Rplot8.png Rplot9.png
## Rplot10.png "beer_ingredients.gif""' execution failed with error code 1
```

```
## Warning: running command '""C:\Program Files\ImageMagick-7.0.4-
## Q16\convert.exe" -loop 0 -delay 100 Rplot1.png Rplot2.png Rplot3.png
## Rplot4.png Rplot5.png Rplot6.png Rplot7.png Rplot8.png Rplot9.png
## Rplot10.png "beer_ingredients.gif""' had status 127
```

```
## an error occurred in the conversion... see Notes in ?im.convert
```

```
## Warning in normalizePath(path.expand(path), winslash, mustWork):
## path[1]="beer_ingredients.gif": The system cannot find the file specified
```

```
## [1] FALSE
```