

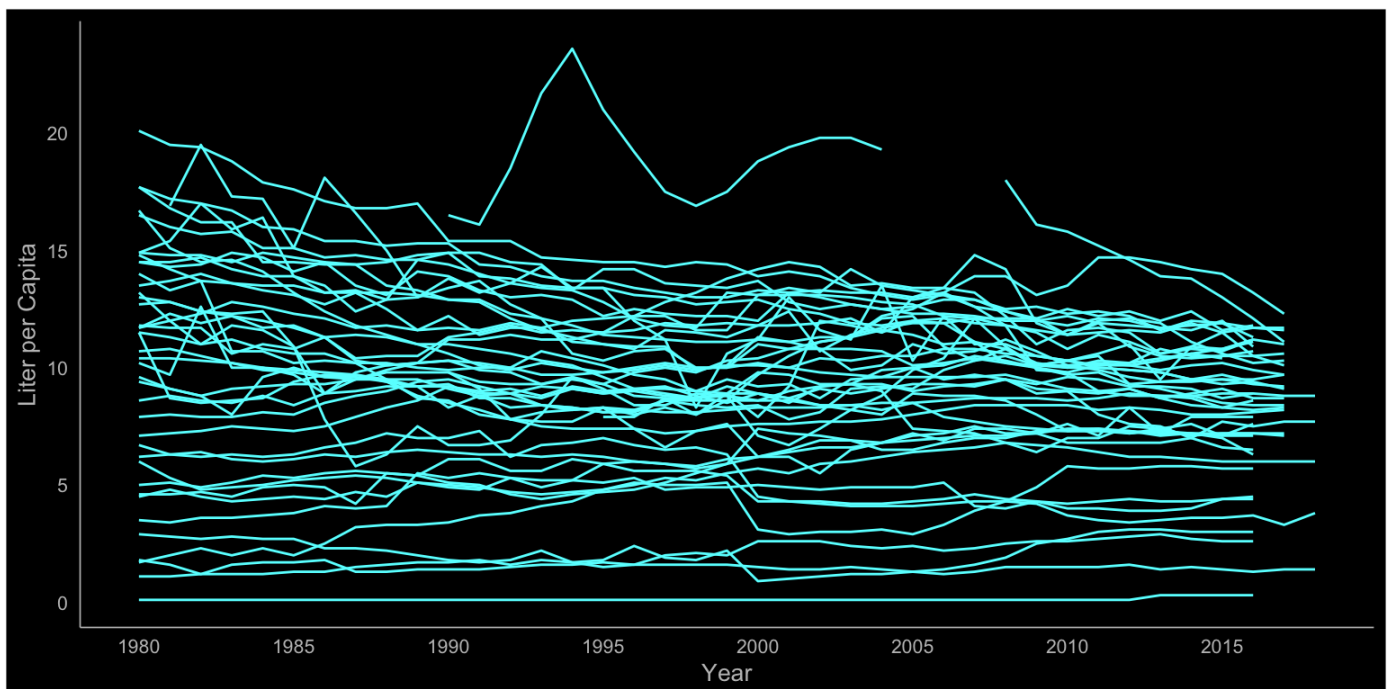
# Data Visualization Group D Deliverable 1

Group D

16-01-2020

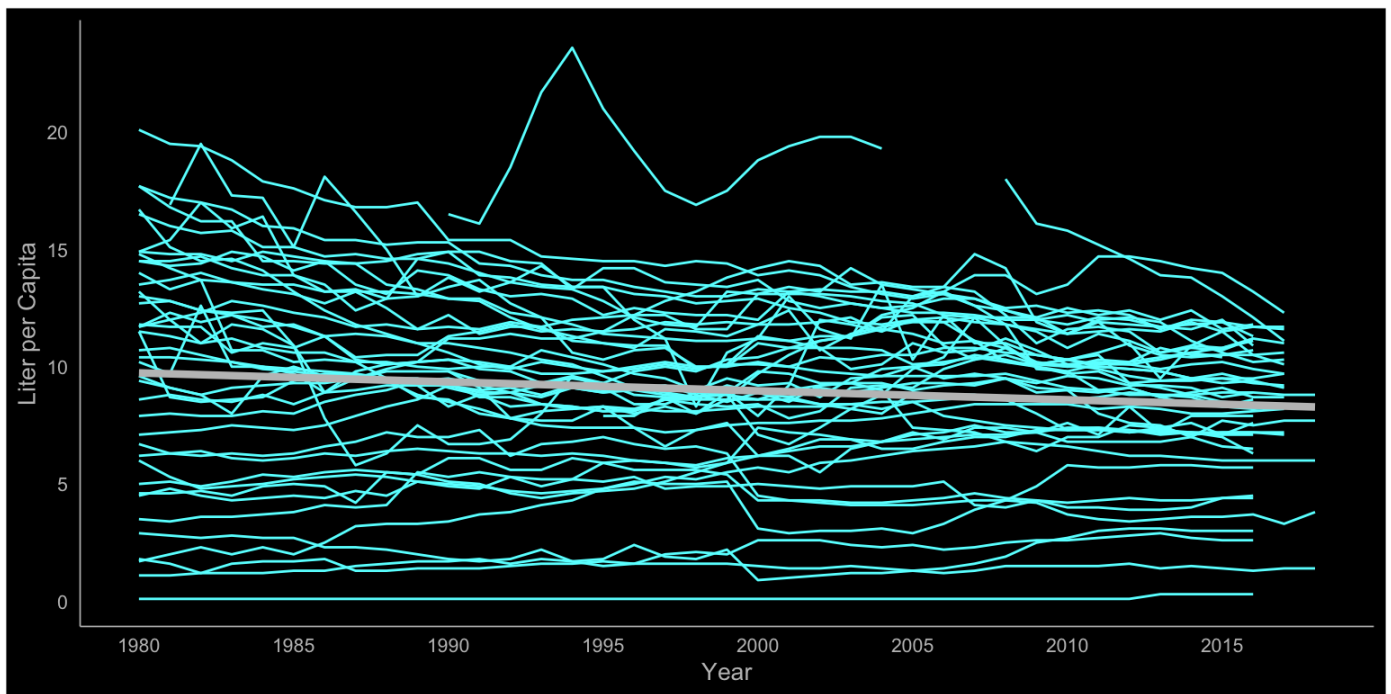
## CHAPTER 1: ALCOHOL CONSUMPTION

```
#Acohol consumption per country
ggplot(Alcohol_Smoking, aes(Year, Literpercapita, group=Country)) +   geom_line(colour = main2_color)+
  labs(x = "Year", y = "Liter per Capita")+
  scale_x_continuous(breaks=seq(1980, 2018,5),labels = seq(1980,2018,5))
```



Alcohol consumption in litres per capita from 1980 - 2018 shows a slight declining trend

```
#Adding a main trendline
ggplot(Alcohol_Smoking, aes(Year, Literpercapita, group=Country)) +   geom_line(colour = main2_color)+
  labs(x = "Year", y = "Liter per Capita")+
  geom_smooth(aes(group = 1), colour = decoration_color, size = 1.5, method = "lm", se = FALSE)+
  scale_x_continuous(breaks=seq(1980, 2018,5),labels = seq(1980,2018,5))
```



Alcohol consumption in litres per capita from 1980 - 2018 with a main trainline shows and confirm a very small decline

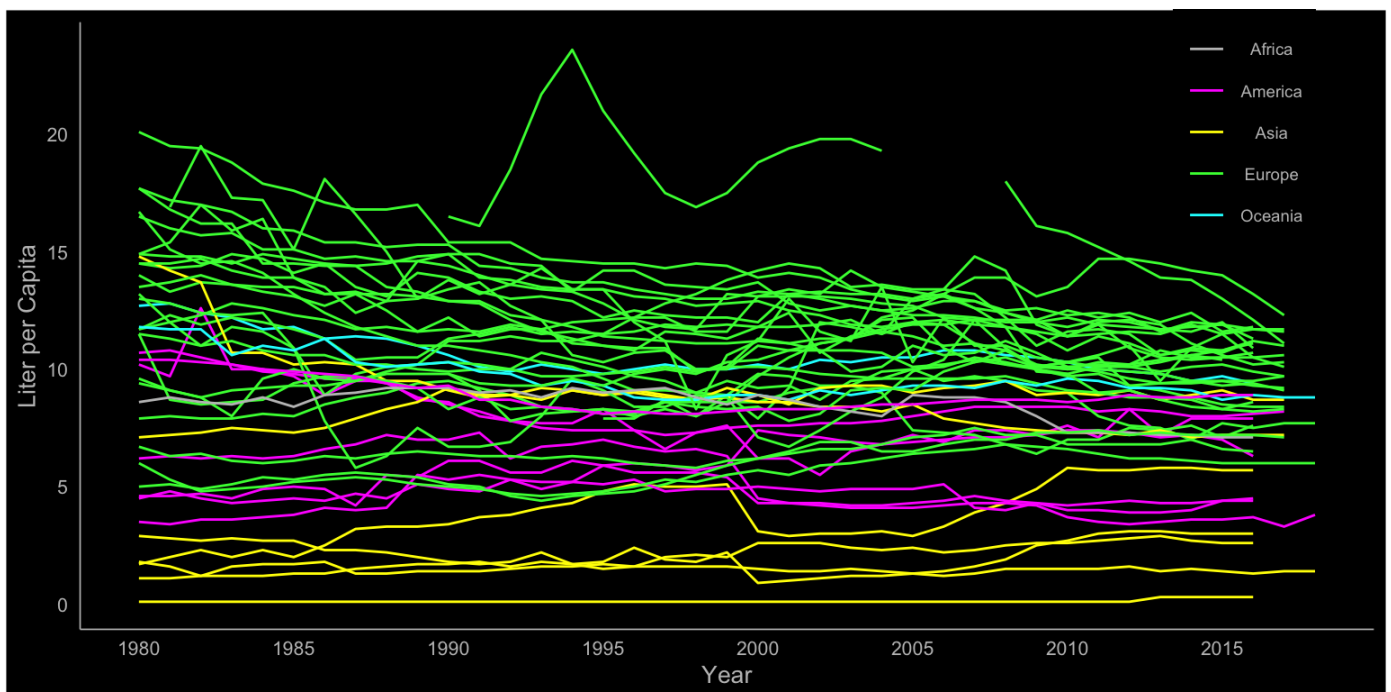
```
#Advance viz. Alcohol liters per Continent
p <- ggplot(Alcohol_Smoking[Literpercapita != "NA",list(Literpercapita= avg(Literperc
apita)), by = c("Continent","Year")], aes(x = Year, y=Literpercapita,group = Continen
t, colour = Continent)) +
  geom_point(aes(x=Year, y=Literpercapita, color= Continent),show.legend = TRUE, alph
a = 0.7, size = 4) +
  labs(x = "Year", y = "Liter per Capita")+
  theme(legend.position = c(0.90, 0.90)) +
  scale_color_manual(values=c("#C0C0C0", "#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))
+
  theme(axis.text.x = element_blank())

p + transition_time(Year) +
  labs(title = "Year: {frame_time}")
```



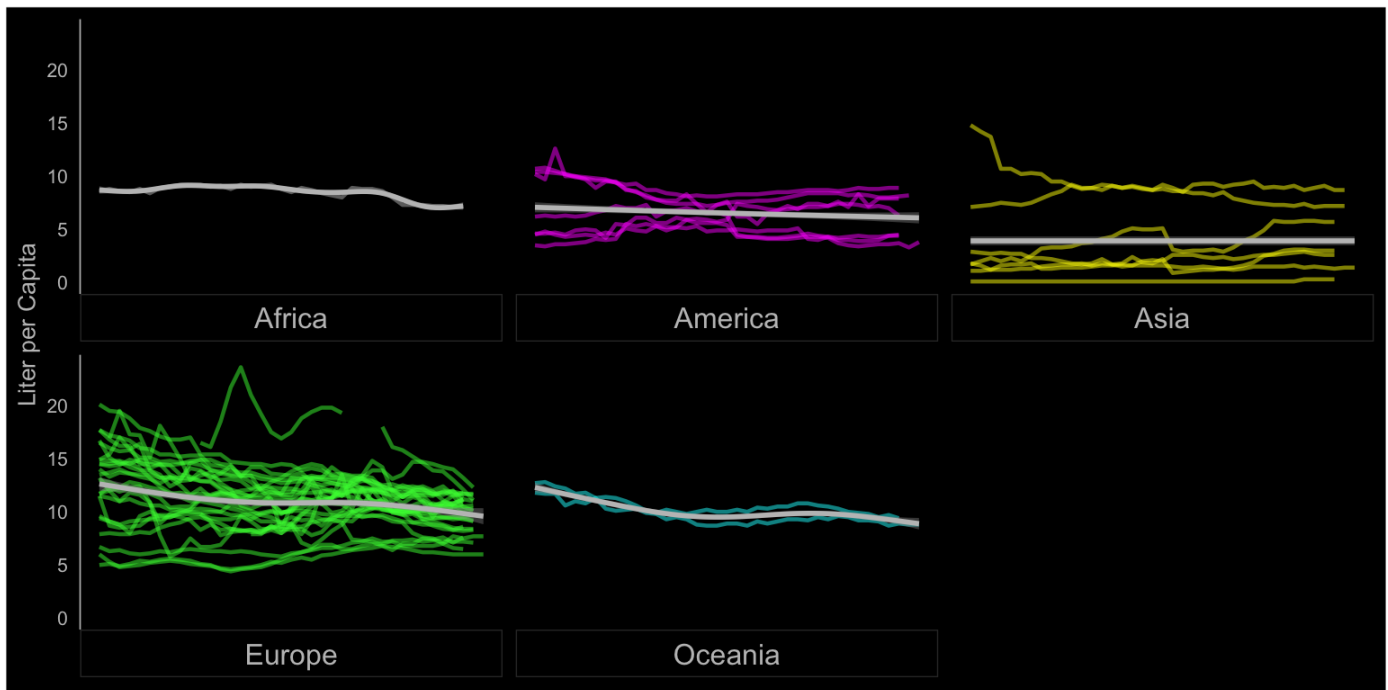
Alcohol consumption in litres per continent seems to converge over time.

```
#Trendlines per country colors by continent
ggplot(Alcohol_Smoking, aes(Year, Literpercapita, group=Country, colour = Continent))
+
  geom_line()+
  labs(x = "Year", y = "Liter per Capita")+
  scale_color_manual(values=c("#C0C0C0", "#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))
+
  theme(legend.position = c(0.90, 0.83), legend.text = element_text(size = 7, hjust
t = 0.5, color = decoration_color))+
  scale_x_continuous(breaks=seq(1980, 2018,5), labels = seq(1980,2018,5))
```



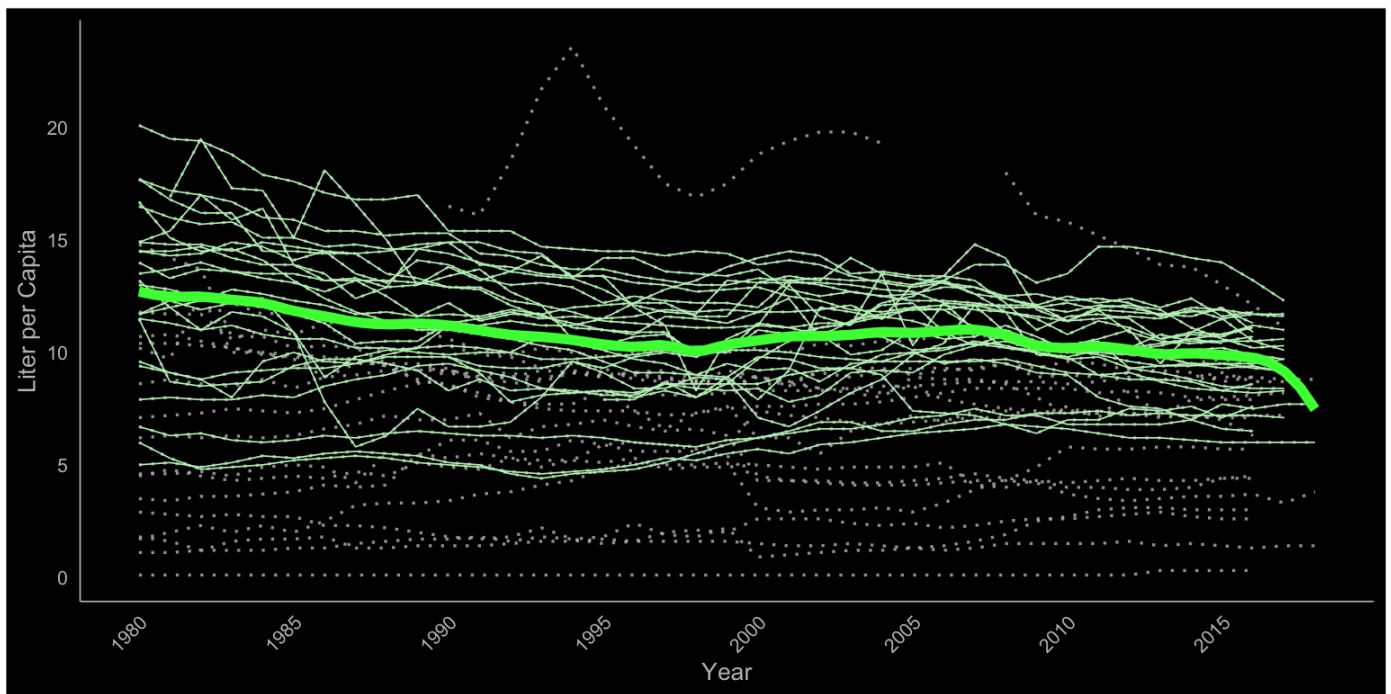
Same graph and comment as above, however showing the distinction of highest and lowest consumption by continent.

```
#Smallmultiple: Trendlines per country colors by continent
ggplot(Alcohol_Smoking, aes(Year, Literpercapita, group=Country,color= Continent)) +
  geom_line(size=0.8, alpha=0.5)+
  labs(x = "Year", y = "Liter per Capita")+
  facet_wrap( ~ Continent, ncol=3, strip.position = "bottom")+
  scale_color_manual(values=c("#C0C0C0", "#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))
+
  stat_smooth(aes(group=1),color=decoration_color)+
  theme(legend.position = "none")+
  theme(axis.text.x = element_blank(), axis.title.x = element_blank(),axis.line.x = element_blank())
```



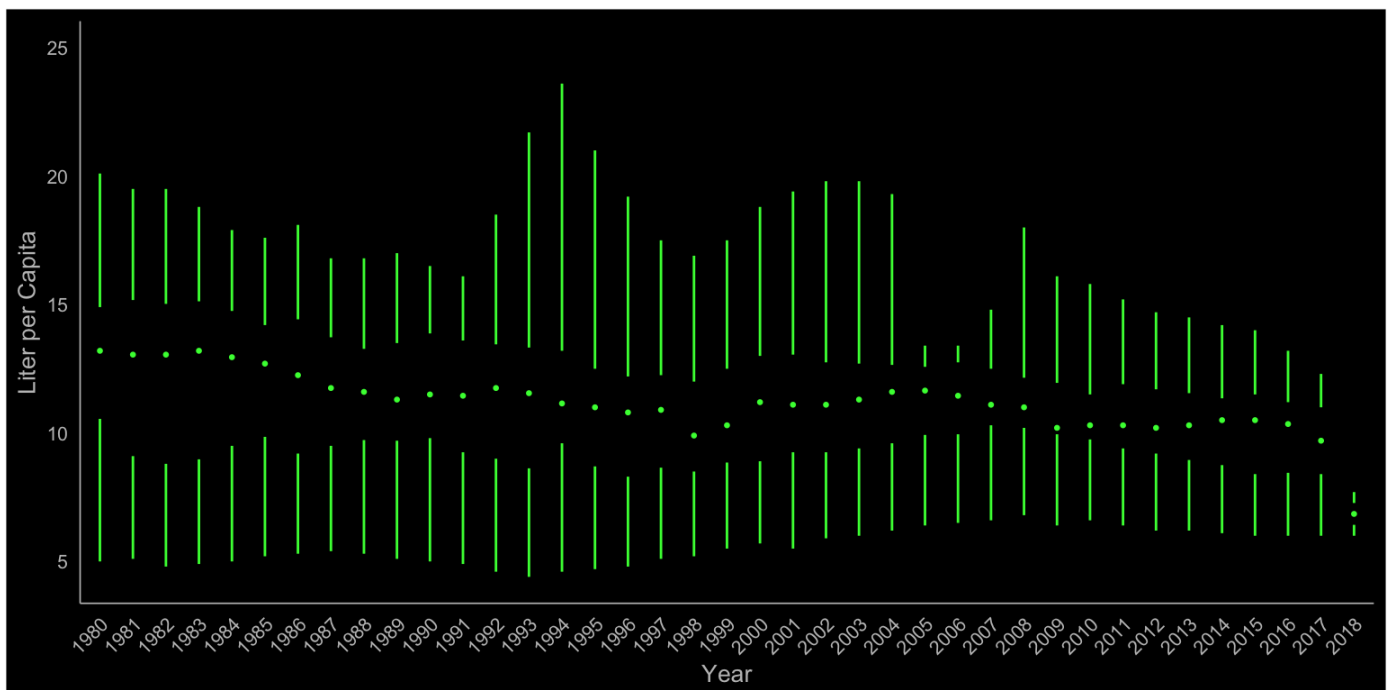
Here we have the same distinction by continent, however with a 5 separate graphs per continent showing individual trend lines. A more clear decreasing trend line can be observed in Europe and Oceania.

```
#All Countries. Trendlines for European countries at the front
ggplot() +
  geom_line(data = transform(Alcohol_Smoking, continent = NULL), aes (Year, Literpercapita, group = Country), alpha = 0.8, lwd = 0.6, colour = decoration_color, linetype = "dotted") +
  geom_line(data=Alcohol_Smoking2[Continent=="Europe"], aes (Year, Literpercapita, group = Country), lwd = 0.3, show.legend = FALSE, color= "#BEF2BF") +
  geom_smooth(data=Alcohol_Smoking2[Continent=="Europe"], aes(Year, Literpercapita, group = 1), lwd = 2, method = 'loess', span = 0.1, se = FALSE, color = "#40FF40") +
  theme(strip.background = element_blank(), strip.placement = "outside") +
  labs(x = "Year", y = "Liter per Capita")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  scale_x_continuous(breaks=seq(1980, 2018,5),labels = seq(1980,2018,5))
```



We decided to highlight the countries in Europe due relevance, high consumption, and high availability of data present in the dataset

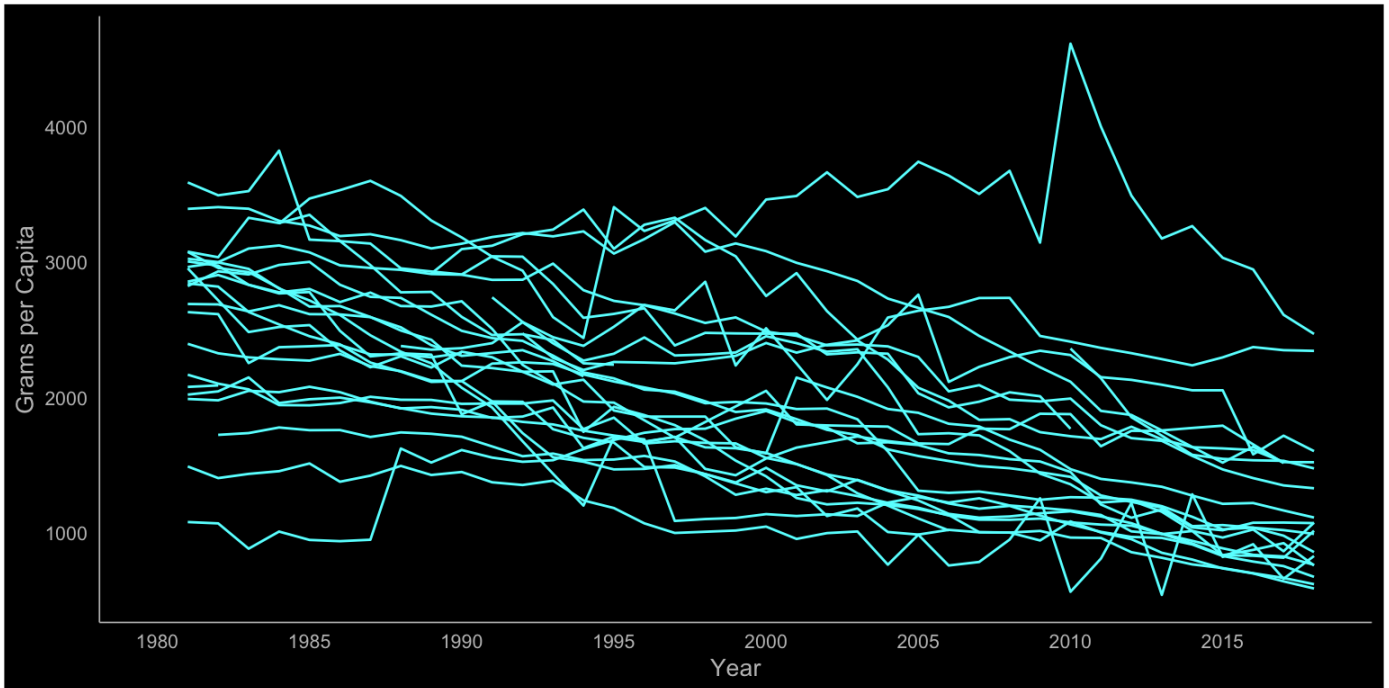
```
#Boxplot European countries
ggplot(Alcohol_Smoking[Continent=="Europe"], aes(factor(Year), Literpercapita)) +
  geom_tufteboxplot(outlier.colour="transparent", color= "#40FF40") +
  annotate("text", x = 10, y = 25, adj=1, family="serif", label = c("")) +
  labs(x = "Year", y = "Liter per Capita")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



The box plot of alcohol consumption in Europe shows how over time there is a decrease in the spread of consumption, whereas the countries who were consuming more are now consuming less and vice versa.

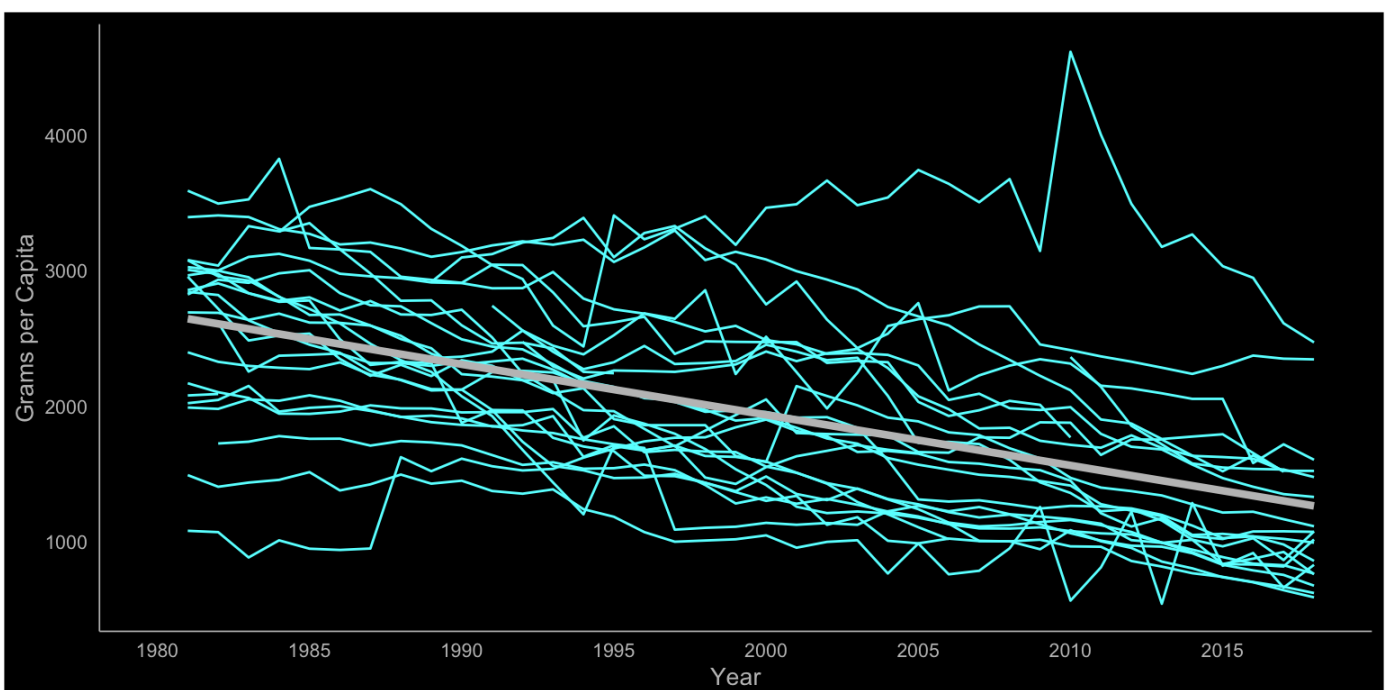
## CHAPTER 2: TOBACCO CONSUMPTION

```
#Trendlines per country
ggplot(Alcohol_Smoking2, aes(Year, Gramspercapita, group=Country))+ geom_line(colour = main2_color)+
  labs(x = "Year", y = "Grams per Capita")+
  scale_x_continuous(breaks=seq(1980, 2018,5),labels = seq(1980,2018,5))
```



Tabacco consumption per country seems to be declining over time.

```
#Adding a main trainline
ggplot(Alcohol_Smoking2, aes(Year, Gramspercapita, group=Country))+ geom_line(colour = main2_color)+
  labs(x = "Year", y = "Grams per Capita")+
  geom_smooth(aes(group = 1), colour = decoration_color, size = 1.5, method = "lm", se = FALSE)+
  scale_x_continuous(breaks=seq(1980, 2018,5),labels = seq(1980,2018,5))
```



Adding a trend line confirms the previous statement of declining worldwide tabacco consumption rates.

```
#Advance viz. Smoking Tobacco grams per Continent
p <- ggplot(Alcohol_Smoking2[Gramspercapita != "NA",list(Gramspercapita= avg(Gramsper
capita)), by = c("Continent","Year")], aes(x = Year, y=Gramspercapita,group = Contine
nt, colour = Continent)) +
  geom_point(aes(x=Year, y=Gramspercapita, color= Continent),show.legend = TRUE, alph
a = 0.7, size = 4) +
  labs(x = "Year", y = "Grams per Capita")+
  theme(legend.position = c(0.90, 0.90)) +
  scale_color_manual(values=c("#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))+
  theme(axis.text.x = element_blank())

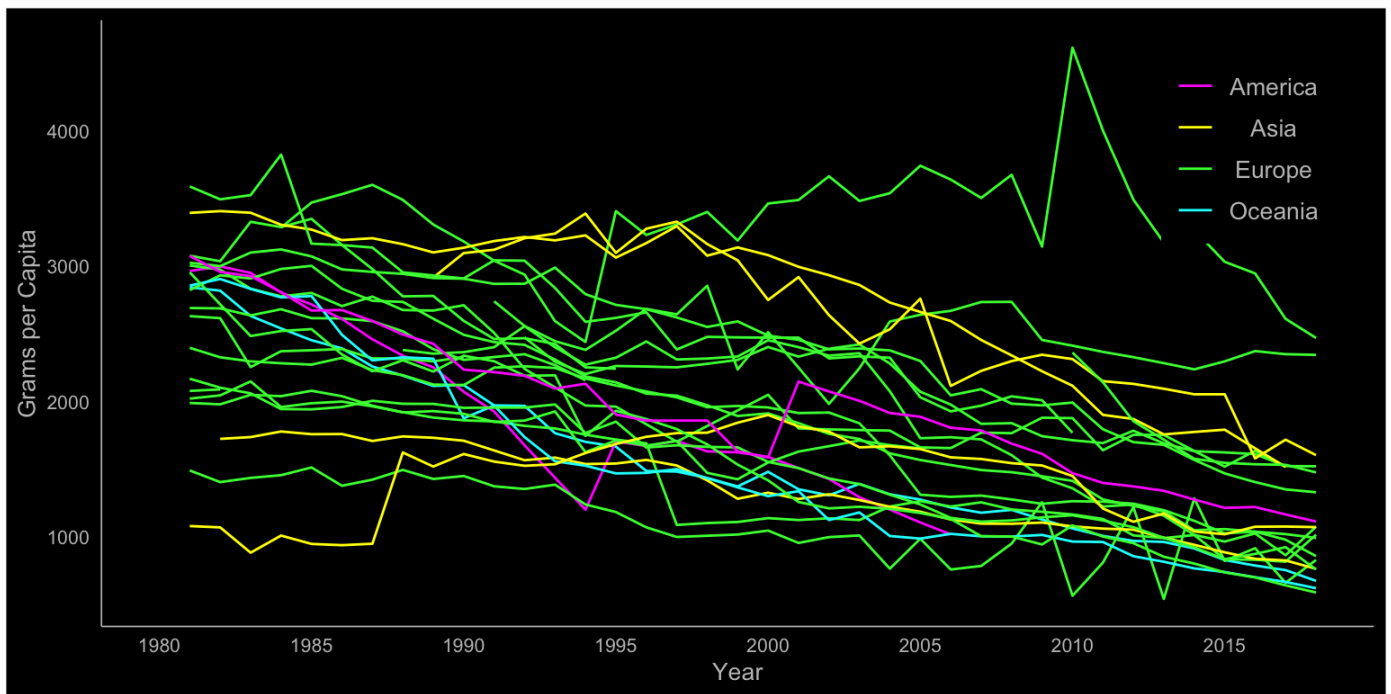
p + transition_time(Year) +
  labs(title = "Year: {frame_time}")
```



The animated version of alcohol consumption per capita again gives us a clear visualization of the declining behavior over time.

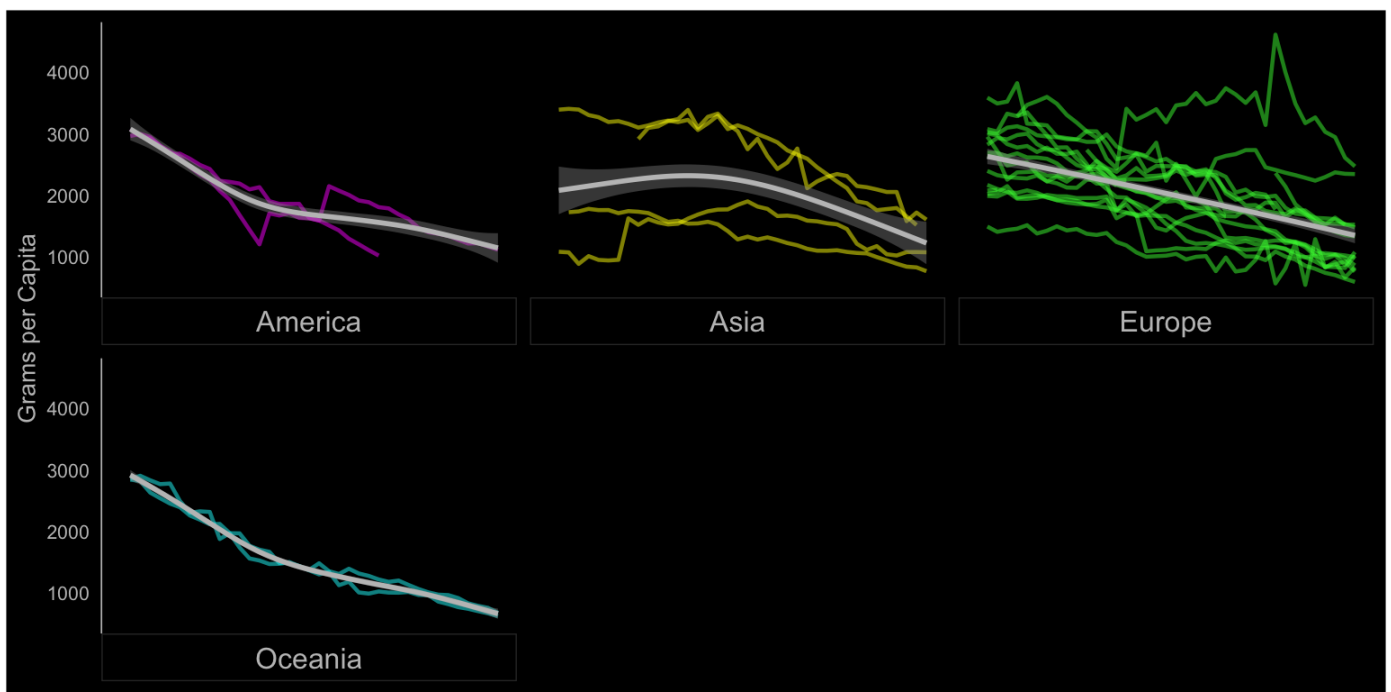
```
#Trendlines per country colors by continent
ggplot(Alcohol_Smoking2, aes(Year, Gramspercapita, group=Country, colour = Continen
t)) +
  geom_line()+
  labs(x = "Year", y = "Grams per Capita")+
  scale_color_manual(values=c("#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))+
  scale_x_continuous(breaks=seq(1980, 2018,5),labels = seq(1980,2018,5))
```





Tabacco consupstion per capita highlighted by contnrent shows us which continents had initally more or less consumption and their decline.

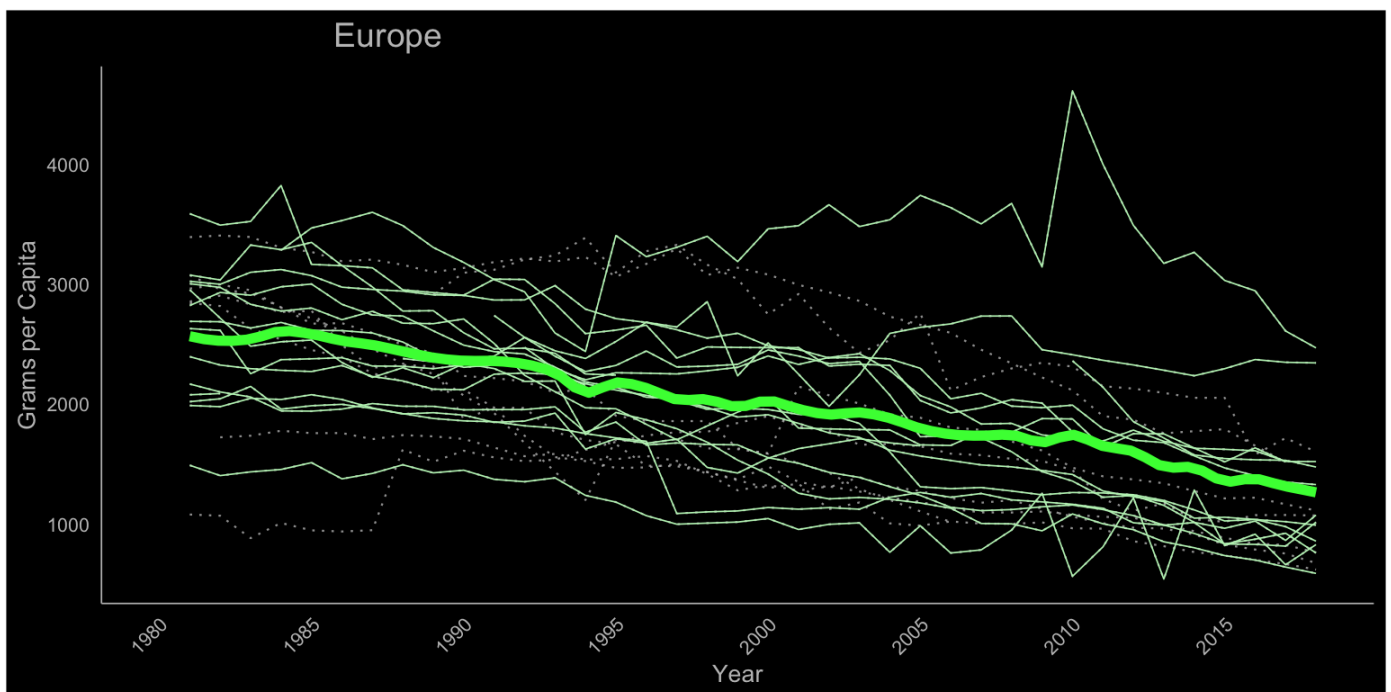
```
#Smallmultiple: Trendlines per country colors by continent
ggplot(Alcohol_Smoking2, aes(Year, Gramspercapita, group=Country,color= Continent)) +
  geom_line(size=0.8, alpha=0.5)+
  labs(x = "Year", y = "Grams per Capita")+
  facet_wrap( ~ Continent, ncol=3, strip.position = "bottom")+
  scale_color_manual(values=c("#FF40FF", "#FFFF00", "#40FF40", "#00FFFF")) +
  stat_smooth(aes(group=1),color=decoration_color)+
  theme(legend.position = "none")+
  theme(axis.text.x = element_blank(), axis.title.x = element_blank(),axis.line.x = e
lement_blank())
```



By sperating tabacco consumption by conitinent in their own individual graphs with treand lines we have a more clear understanding of their respective declining be havior, espeacialy in Europe.

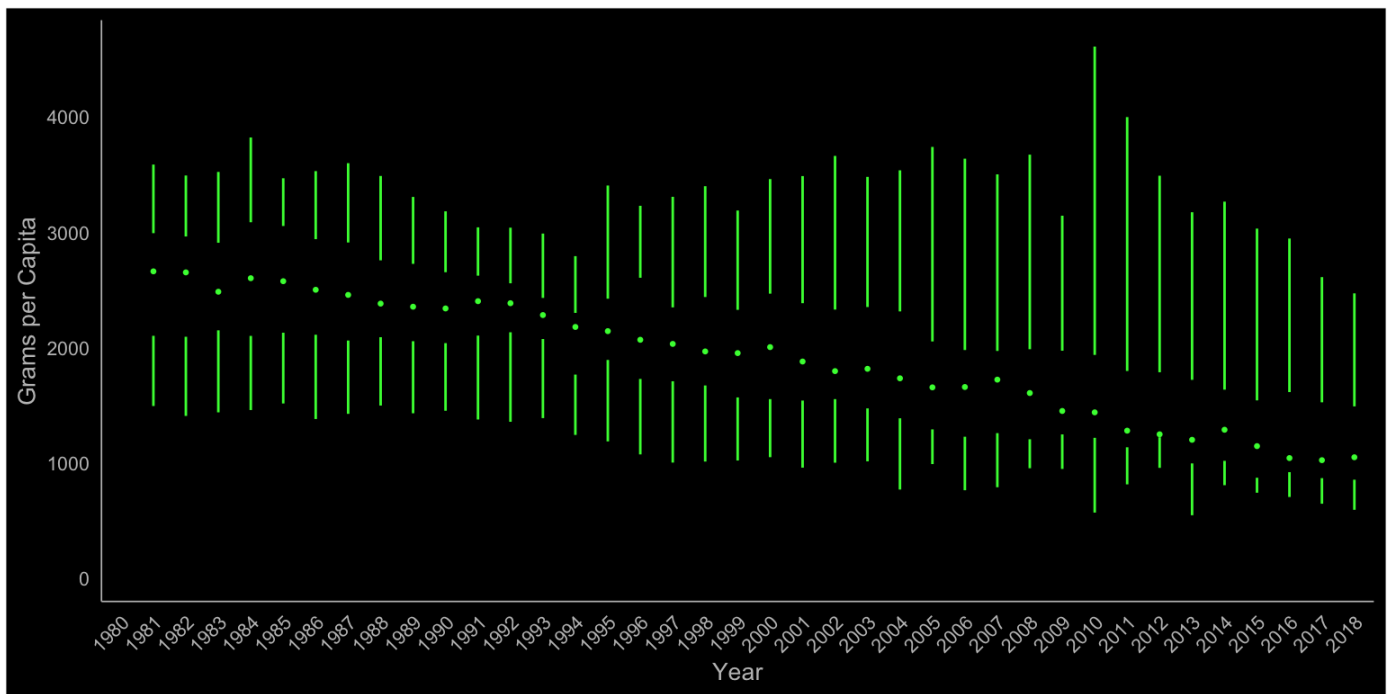


```
#All data. Trendlines for Europe per country
ggplot() +
  geom_line(data = transform(Alcohol_Smoking2, continent = NULL), aes (Year, Gramsper
capita, group = Country), alpha = 0.8, lwd = 0.4, colour = decoration_color, linetype
= "dotted") +
  geom_line(data=Alcohol_Smoking2[Continent=="Europe"], aes (Year, Gramspercapita, gr
oup = Country), lwd = 0.3, show.legend = FALSE, color= "#BEF2BF") +
  geom_smooth(data=Alcohol_Smoking2[Continent=="Europe"], aes(Year, Gramspercapita, g
roup = 1), lwd = 2, method = 'loess', span = 0.1, se = FALSE, color = "#40FF40") +
  theme(strip.background = element_blank(), strip.placement = "outside") +
  labs(x = "Year", y = "Grams per Capita")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))+
  labs(title = "Europe")+
  scale_x_continuous(breaks=seq(1980, 2018,5),labels = seq(1980,2018,5))
```



Tabacco consumption with all countries highlighted for Europe with a trend line was chosen due to available data and the previous choice of focusing on Europe for alcohol consumption.

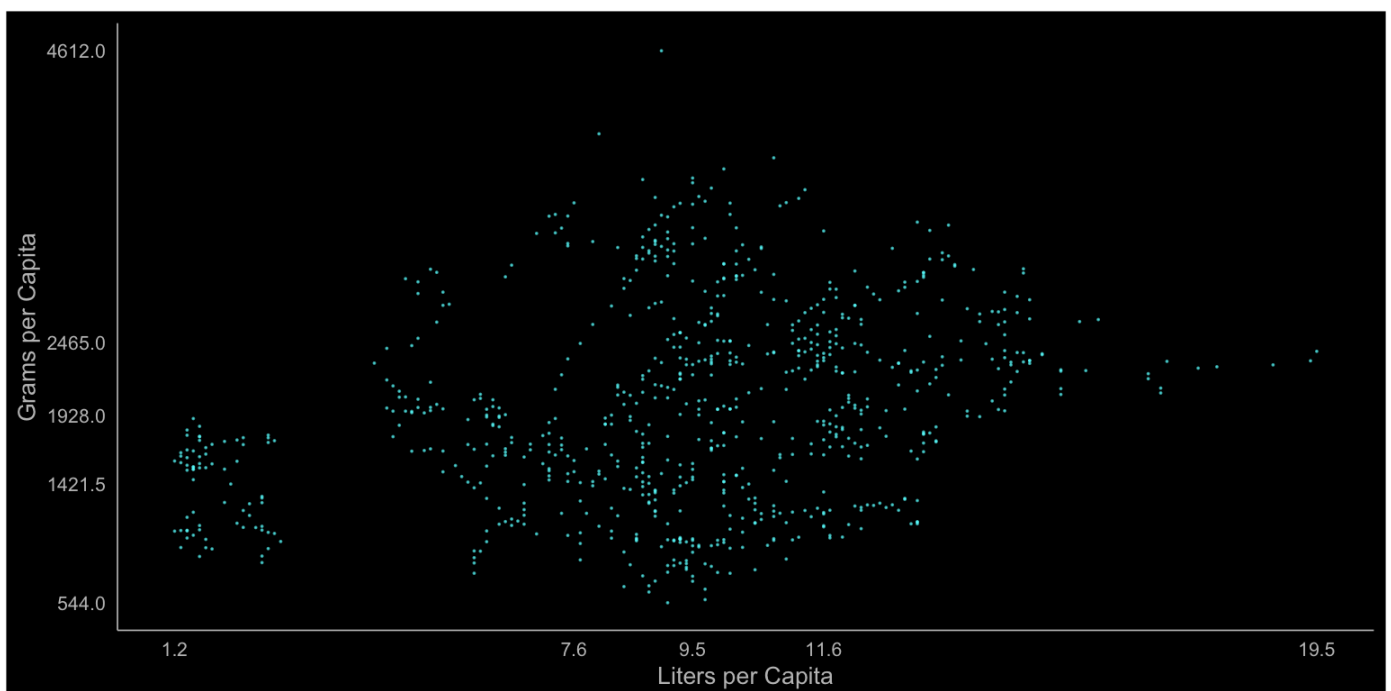
```
#Boxplot European countries
ggplot(Alcohol_Smoking2[Continent=="Europe"], aes(factor(Year),Gramspercapita)) +
  geom_tufteboxplot(outlier.colour="transparent", color= "#40FF40") +
  annotate("text", x = 10, y = 25, adj=1, family="serif", label = c("")) +
  labs(x = "Year", y = "Grams per Capita")+
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



The boxplot for European countries shows the change in dispersion and overall declining rate of tobacco consumption.

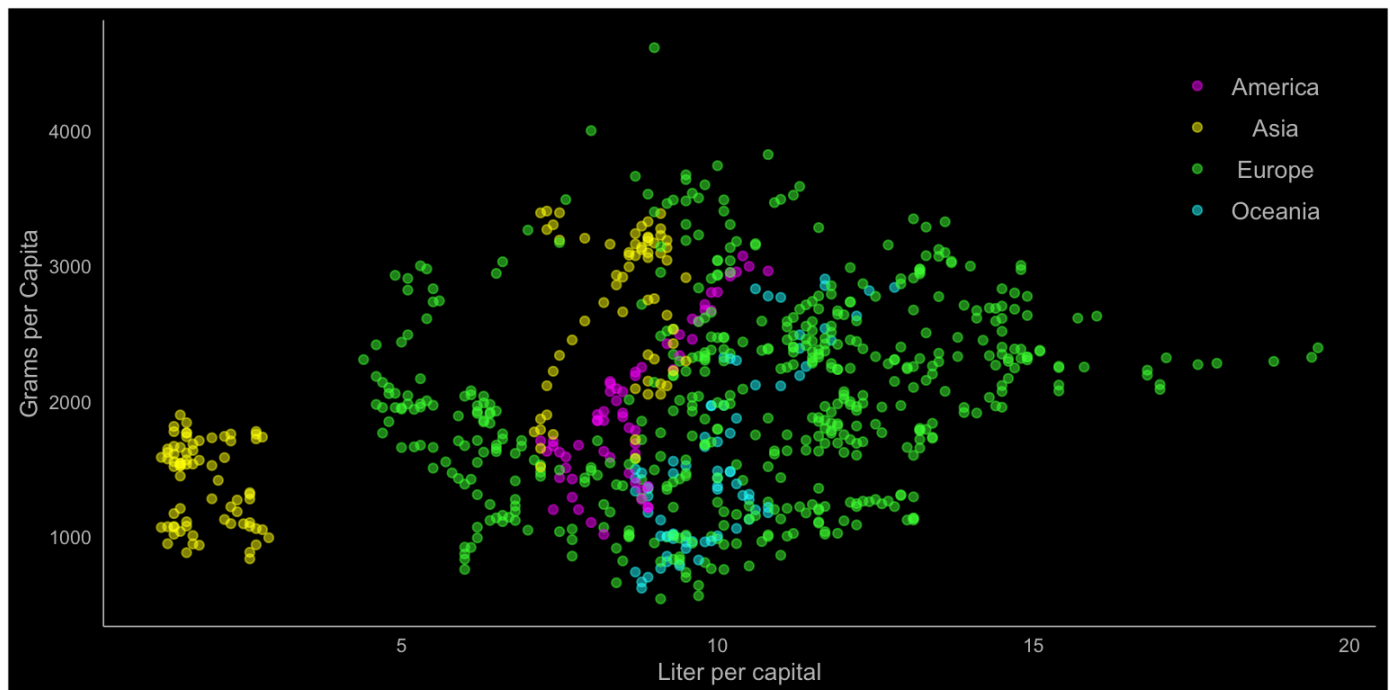
## CHAPTER 3: ALCOHOL AND TOBACCO CONSUMPTION

```
#First we make a basic scatter plot looking for linearity
ggplot(Alcohol_Smoking2, aes(Literpercapita, Gramspercapita)) +
  geom_point(size=0.02, alpha=0.6, color=main2_color) +
  xlab("Liters per Capita")+
  ylab("Grams per Capita")+
  scale_x_continuous(breaks = round(as.vector(quantile(Alcohol_Smoking2$Literpercapita)), digits = 1))+
  scale_y_continuous(breaks = round(as.vector(quantile(Alcohol_Smoking2$Gramspercapita)), digits = 1))
```



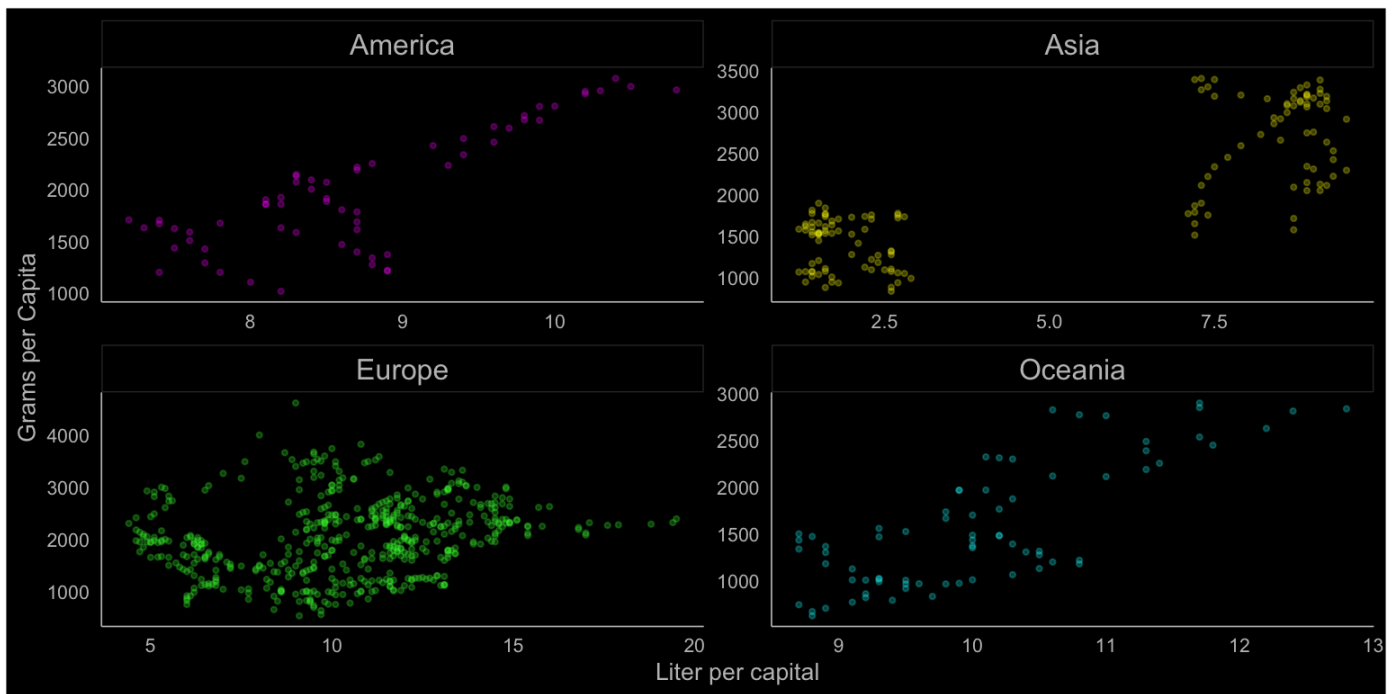
The scatter plot for grams per capita and liter per capita shows no direct relationship, contrary to a logical assumption of a direct positive relationship.

```
#Scatterplot using colors for continent
ggplot(Alcohol_Smoking2, aes(Literpercapita, Gramspercapita, color = Continent)) +
  geom_point(alpha=0.5) +
  labs(x = "Liter per capital", y = "Grams per Capita")+
  scale_color_manual(values=c("#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))
```



From the below illustration, for each continent we still can not see a clear linear association between the two variables

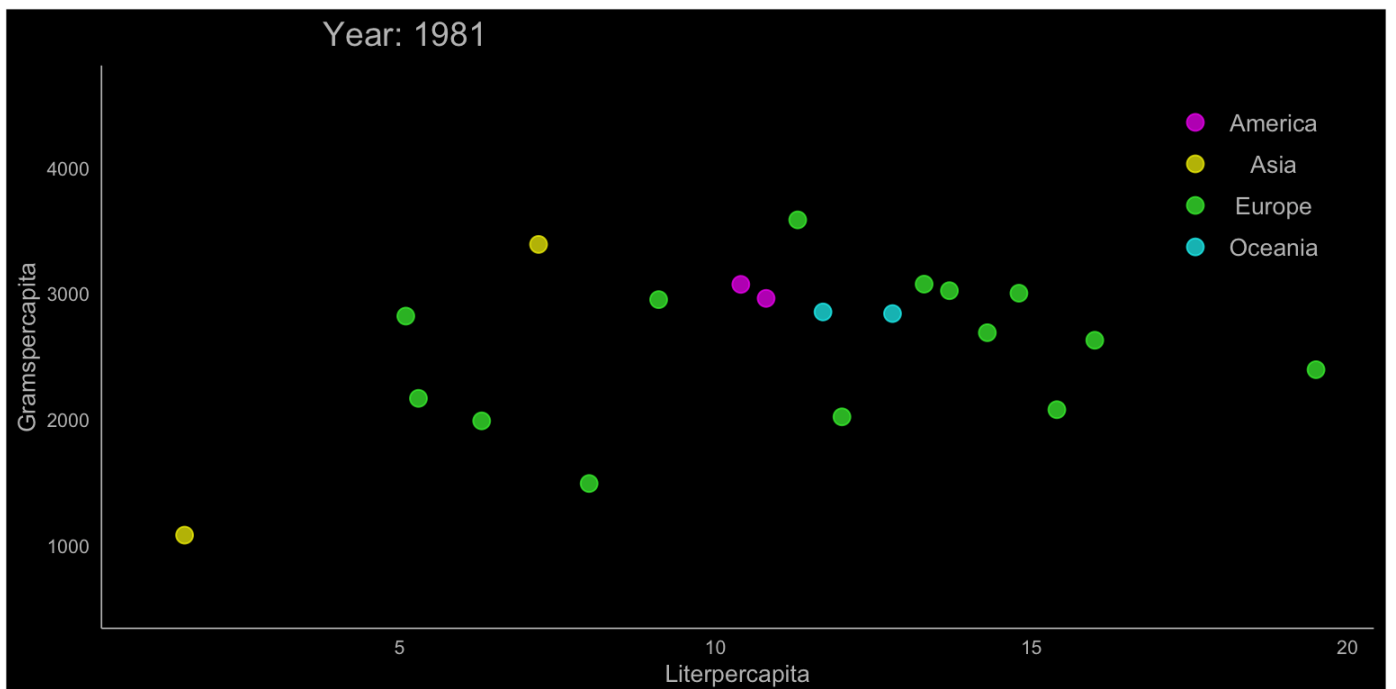
```
#To double check that relationship we use the Smallmultiple Scatterplot to see the linear relationship for each Continent
ggplot(Alcohol_Smoking2, aes(x=Literpercapita, y=Gramspercapita, color = Continent))
+
  geom_point(size=0.8, alpha=0.25)+
  scale_color_manual(values=c("#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))+
  facet_wrap( ~ Continent, ncol=2, scales = "free")+
  labs(x = "Liter per capital", y = "Grams per Capita")+
  theme(legend.position = "none")
```



This scatter plot for grams per capita and liters per capita shows a relationship by continent, and how countries in the America's and Oceania have the greatest positive relationship.

```
#Advance viz. Scatter Plot Evolution
p <- ggplot(Alcohol_Smoking2, aes(Literpercapita, Gramspercapita, color = Continent))
+
  geom_point(alpha = 0.7, size = 3) +
  scale_color_manual(values=c("#FF40FF", "#FFFF00", "#40FF40", "#00FFFF"))

p + transition_time(Year) +
  labs(title = "Year: {frame_time}")
```



This animation of liters per capita and grams of tobacco by per capita over time shows how in general, and despite assumed positive relationship, people are smoking and drinking less.

## General Findings:

1. The average alcohol consumption has been consistently decreasing in the OCDE countries, this is mainly because in Europe (the one that has more countries) has been consistently decreasing. Also the distribution of the alcohol consumptions has narrowed through the years
2. Like Alcohol, Tobacco has seen a sharp decrease in the amount of tobacco consumption thanks to Europe. The main difference in this case is that some countries have started to become bigger outliers and decreased their consumption far less than the average (bigger outliers)
3. As both consumptions have decreased through time, we can see a stronger decrease in tobacco compared to alcohol but after looking at the dynamic scatter plot we can see a consistent behaviour toward decreasing both alcohol and tobacco together