

Intro GGplot

DORE Martin

2023-10-19

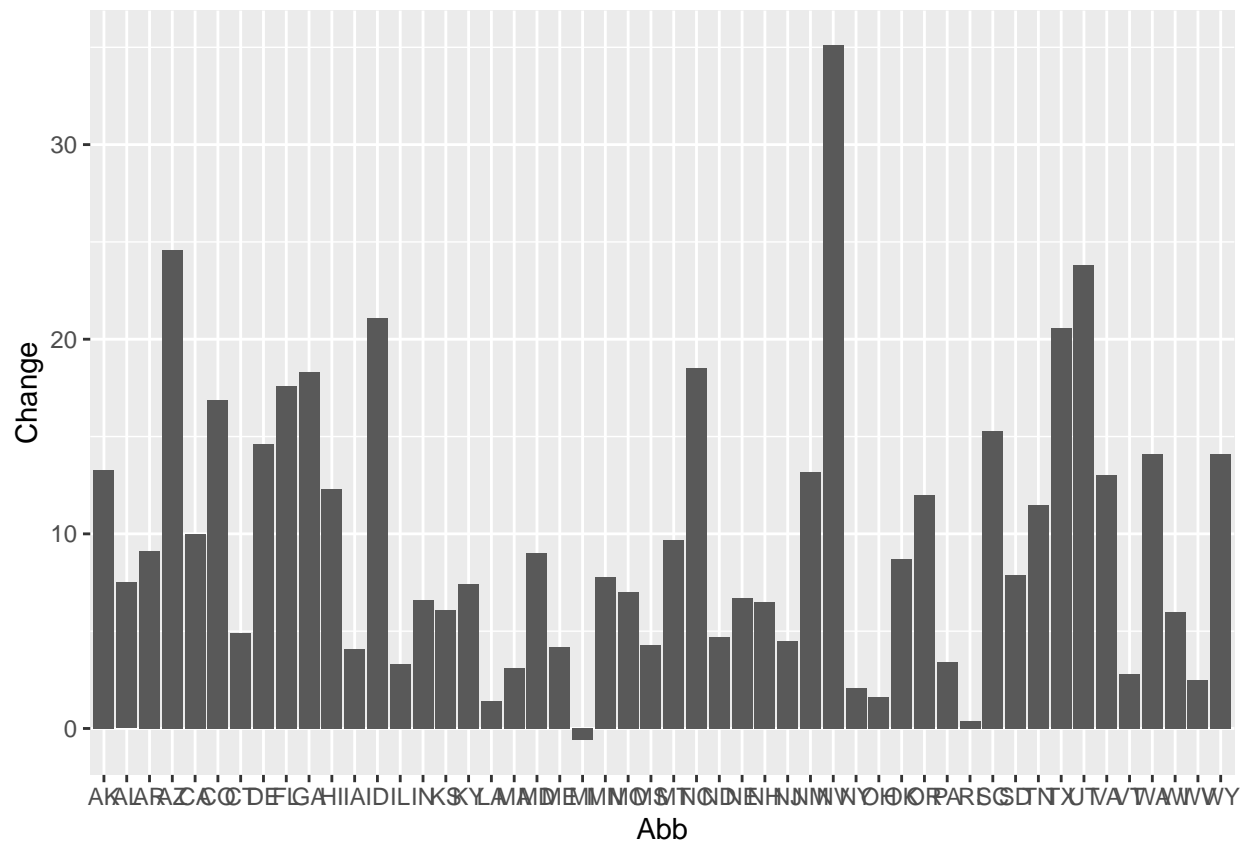
Librairies

```
library(ggplot2)
library(gcookbook)
library(RColorBrewer)
library(paletteer)
library(MASS)
library(grid)
library(dplyr)
set.seed(10)

library(wordcloud2)
uspop10 = uspopchange[sample(nrow(uspopchange), 10),]
```

Barplot

```
plot1 = ggplot(data = uspopchange, aes(x = Abb, y = Change))+
  geom_col()
plot1
```



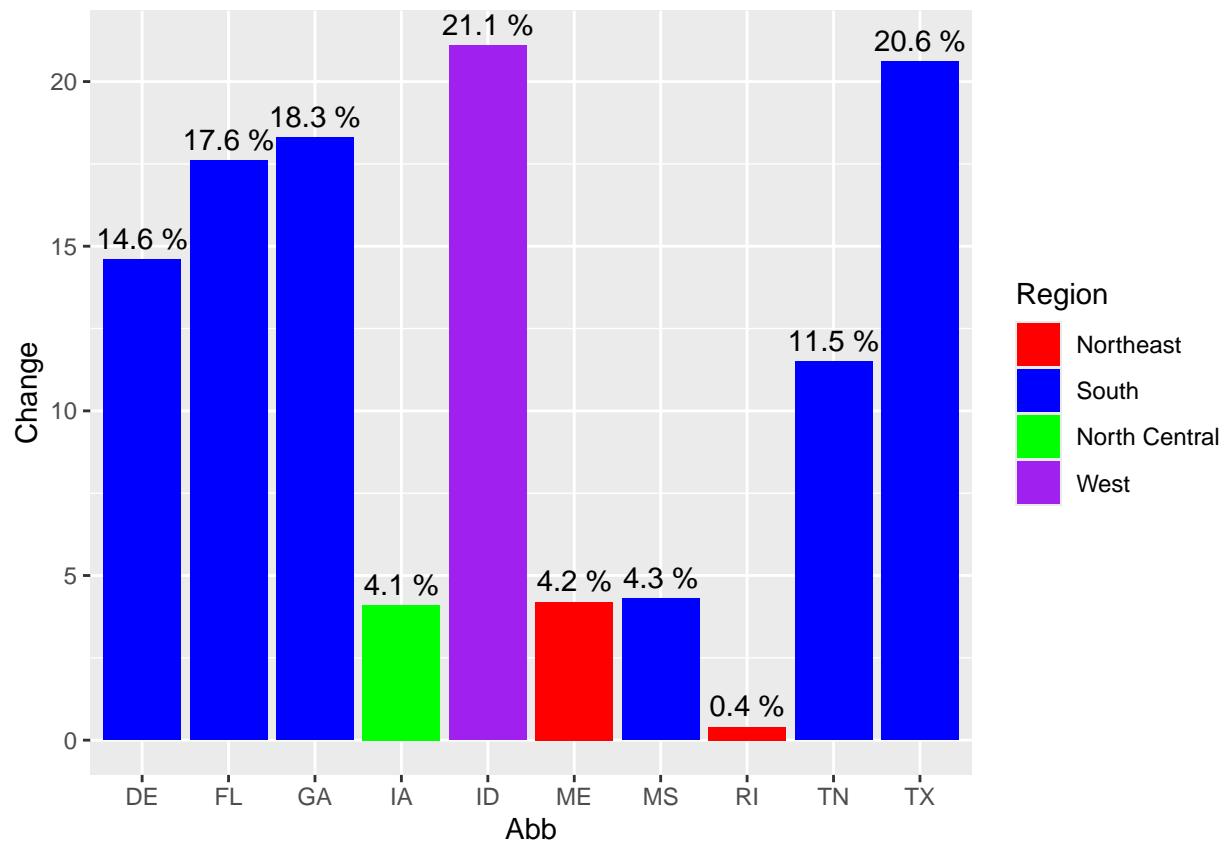
```

uspop10 <- uspop10 %>% arrange(Change)
# Créez un vecteur de couleurs personnalisées pour chaque région
couleurs_personnalisees <- c("Red", "Blue", "Green", "Purple")

# Créez un histogramme en utilisant ggplot2 et spécifiez les couleurs
plot2 = ggplot(data = uspop10, aes(x = Abb, y = Change, fill = Region)) +
  geom_col() +
  scale_fill_manual(values = couleurs_personnalisees) +
  geom_text(aes(label = paste(Change,"%")), vjust = -0.5, color = "black")

# Affichez l'historgramme
plot2

```



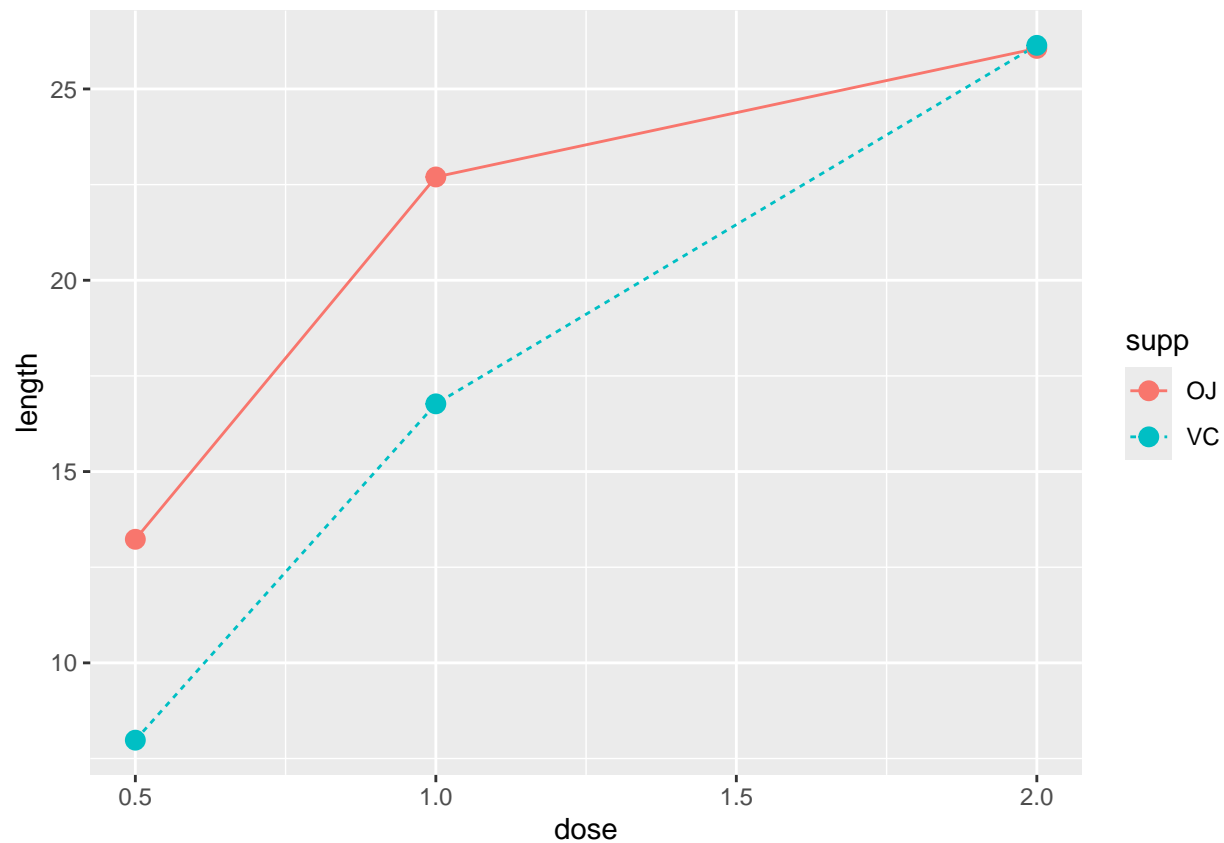
graficos de lineas

```
head(tg)
```

```
##      supp dose length
## 1    OJ  0.5  13.23
## 2    OJ  1.0  22.70
## 3    OJ  2.0  26.06
## 4    VC  0.5   7.98
## 5    VC  1.0  16.77
## 6    VC  2.0  26.14
```

```
p = ggplot(data=tg, aes(x=dose, y = length, fill = supp, color = supp, lty = supp))+
  geom_line()+
  geom_point(shape = 21, size = 3)
```

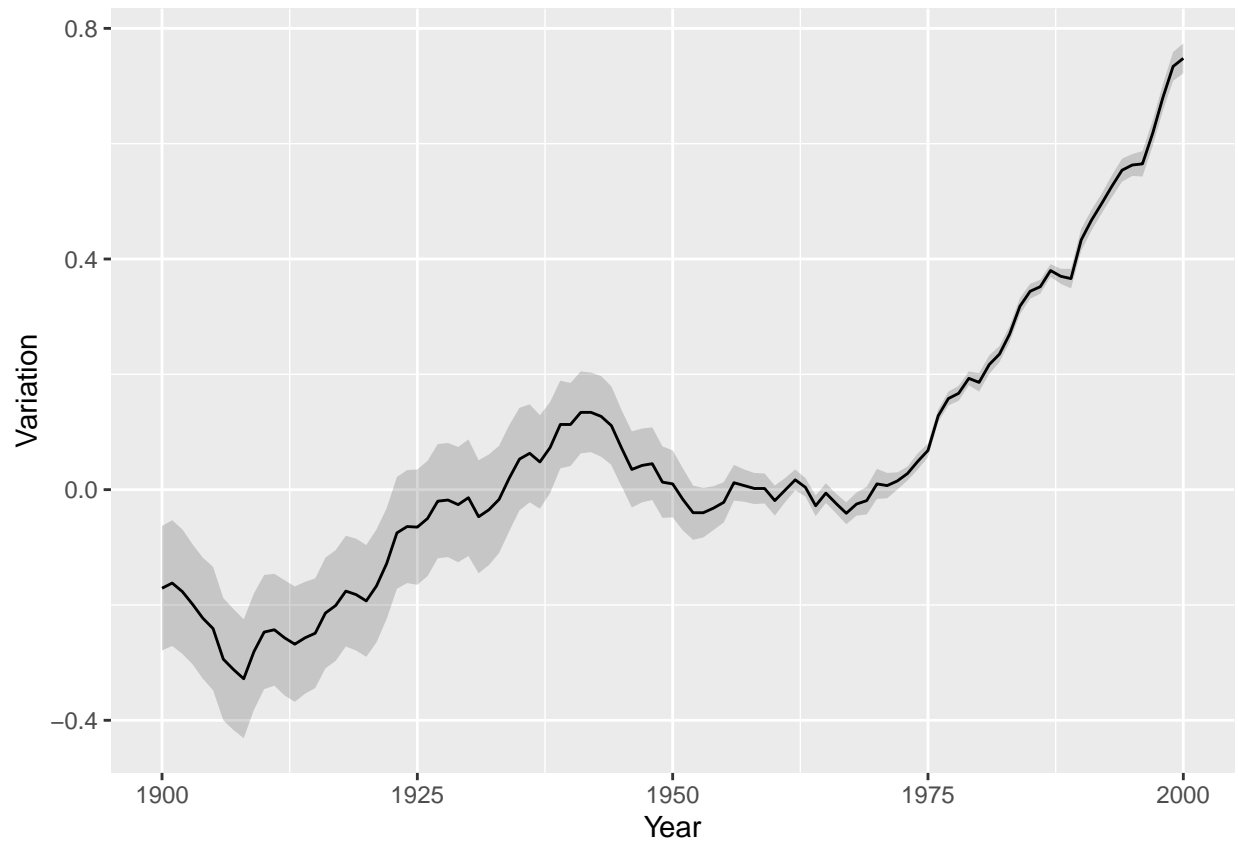
```
p
```



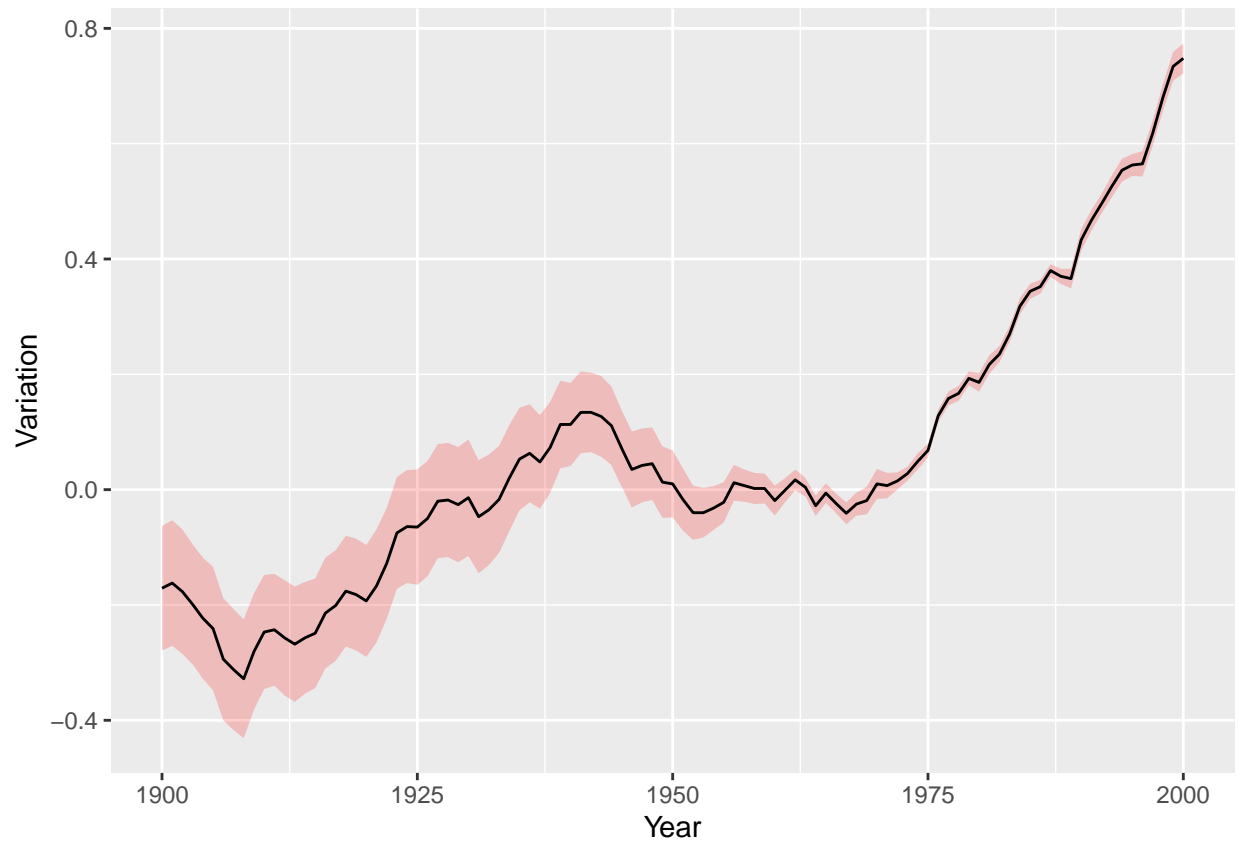
```
library(readr)
Temperatures <- read_csv("Temperatures.csv")
head(Temperatures)
```

```
## # A tibble: 6 x 4
##   Year Variation conf.level95 Positive
##   <dbl>     <dbl>      <dbl> <lgl>
## 1  1900    -0.171      0.108 FALSE
## 2  1901    -0.162      0.109 FALSE
## 3  1902    -0.177      0.108 FALSE
## 4  1903    -0.199      0.104 FALSE
## 5  1904    -0.223      0.105 FALSE
## 6  1905    -0.241      0.107 FALSE
```

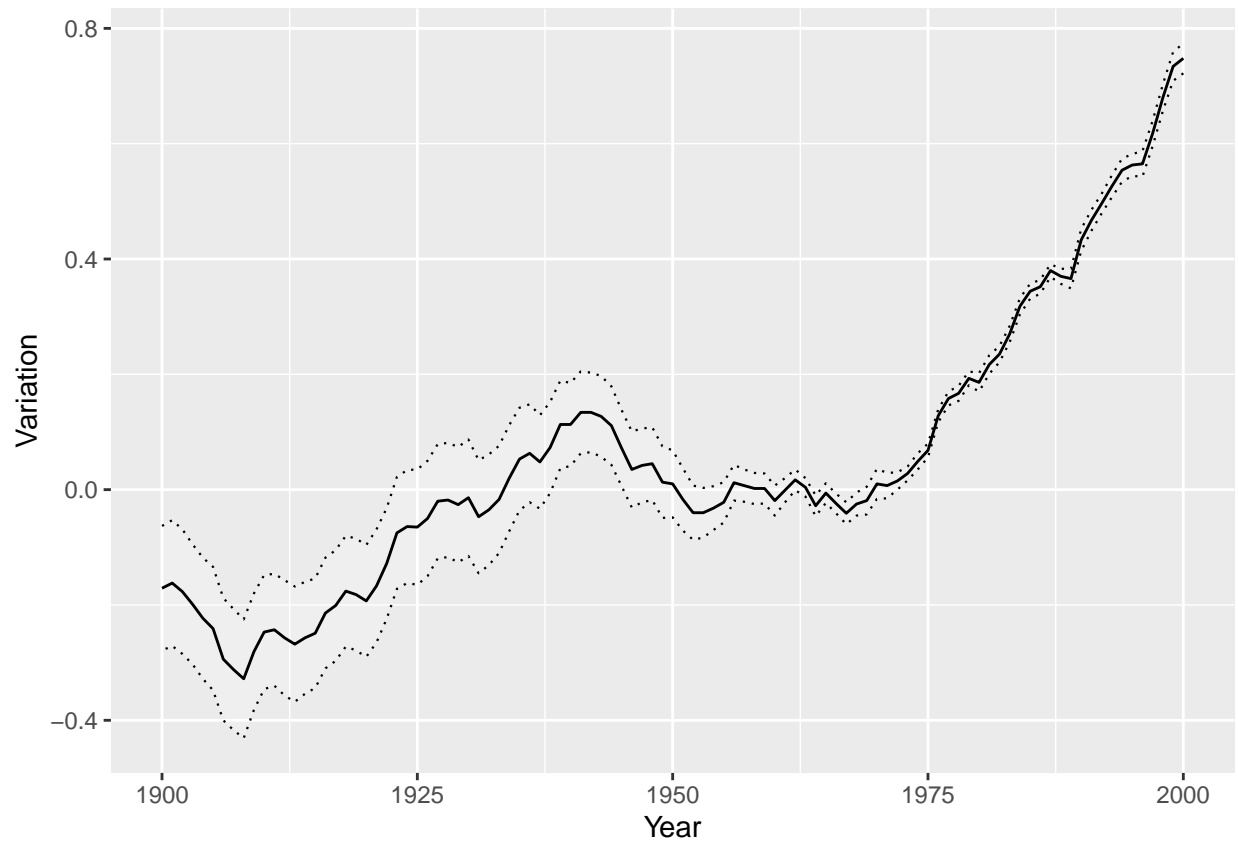
```
ggplot(Temperatures, aes(x = Year, y = Variation)) +
  geom_ribbon(aes(ymin = Variation - conf.level95,
                ymax = Variation + conf.level95),
            alpha = 0.2) +
  geom_line()
```



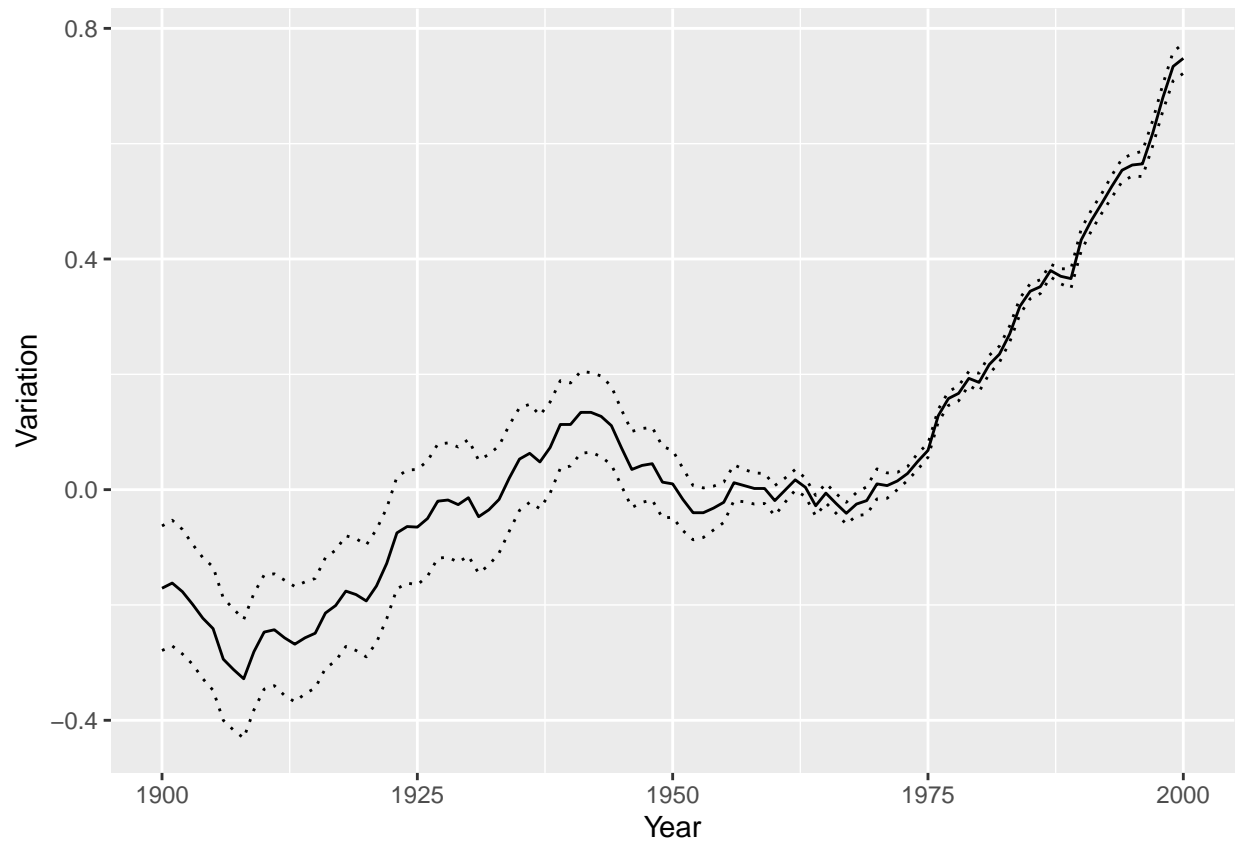
```
ggplot(Temperatures, aes(x = Year, y = Variation)) +  
  geom_ribbon(aes(ymin = Variation - conf.level95,  
                ymax = Variation + conf.level95),  
            alpha = 0.2,  
            fill = "red") +  
  geom_line()
```



```
ggplot(Temperatures, aes(x = Year, y = Variation)) +  
  geom_ribbon(aes(ymin = Variation - conf.level95,  
                ymax = Variation + conf.level95),  
            alpha = 0.2,  
            colour = "black",  
            lty = "dotted",  
            size = 0.4,  
            fill = "white") +  
  geom_line()
```



```
ggplot(Temperatures, aes(x = Year, y = Variation)) +
  geom_line() +
  geom_line(data = Temperatures,
            aes(x = Year, y = Variation - conf.level95),
            linetype = "dotted", color = "black") +
  geom_line(data = Temperatures,
            aes(x = Year, y = Variation + conf.level95),
            linetype = "dotted", color = "black") +
  scale_linetype_manual(values = c("solid", "dotted"))
```



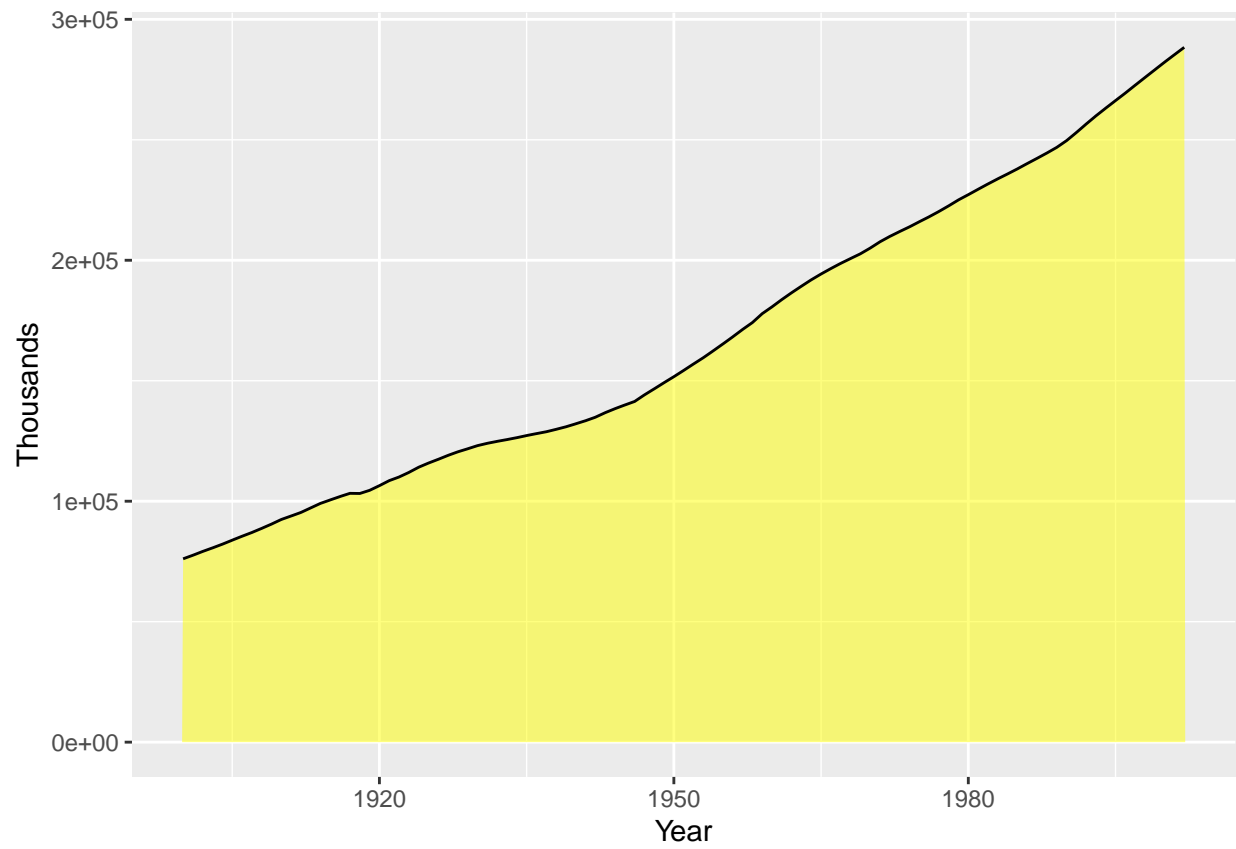
```
head(uspophage, 5)
```

```
##   Year AgeGroup Thousands
## 1 1900      <5      9181
## 2 1900    5-14     16966
## 3 1900   15-24     14951
## 4 1900   25-34     12161
## 5 1900   35-44      9273
```

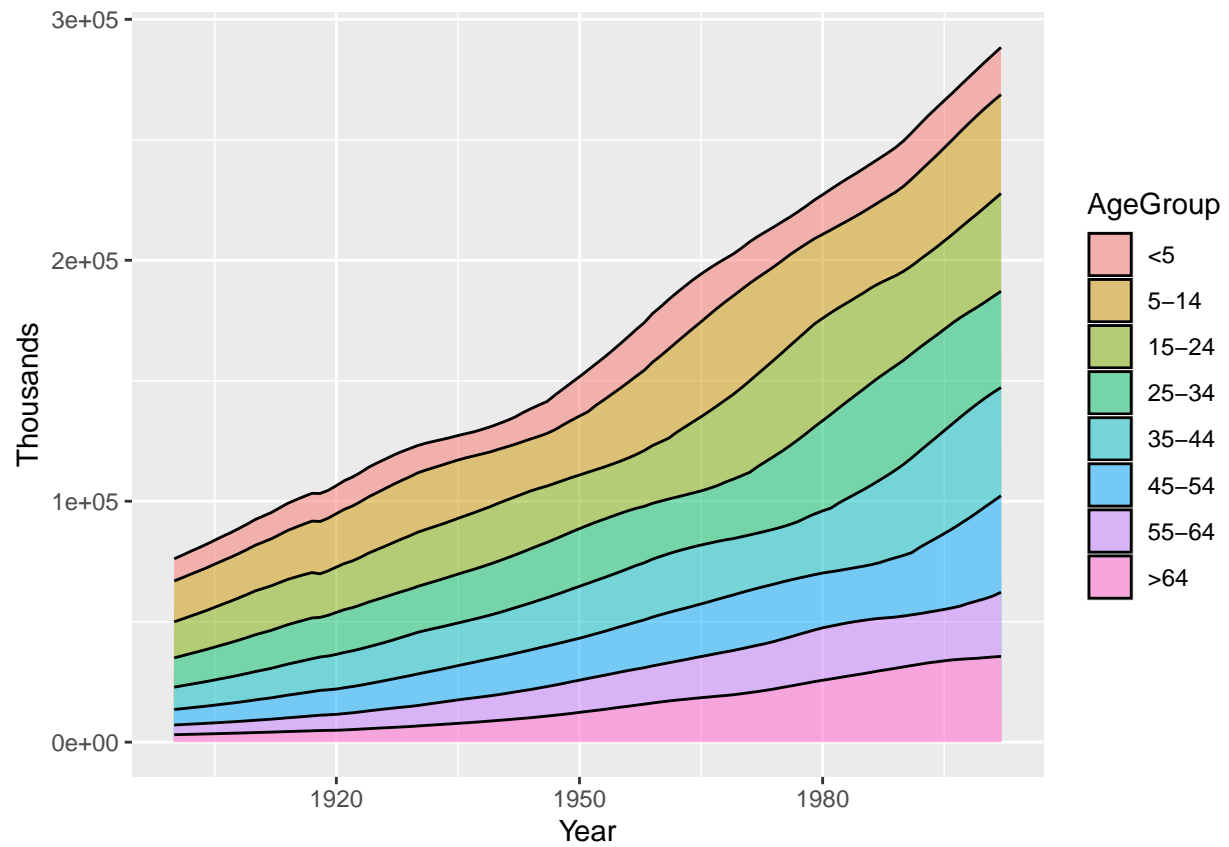
```
# compile les données par années
uspophage_years = aggregate(Thousands ~ Year, data = uspophage, sum)
head(uspophage_years)
```

```
##   Year Thousands
## 1 1900      76094
## 2 1901      77584
## 3 1902      79163
## 4 1903      80632
## 5 1904      82166
## 6 1905      83822
```

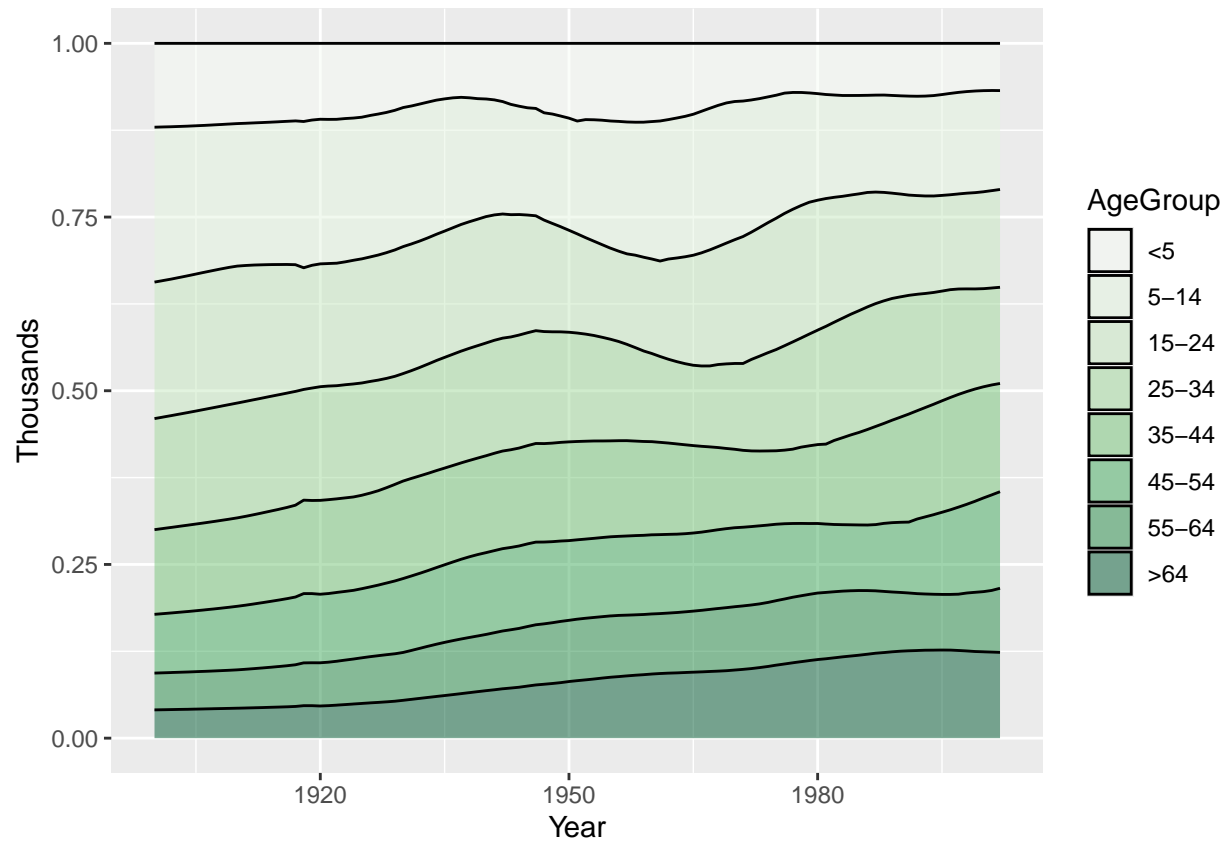
```
ggplot(data = uspophage_years, aes(x = Year, y = Thousands)) +
  geom_area(color = "black", fill = "yellow", alpha = 0.5)
```

```
ggplot(data = uspopage, aes(x = Year, y = Thousands, fill = AgeGroup))+  
  geom_area(color = "black", alpha = .5)
```



```
ggplot(data = uspopage, aes(x = Year, y = Thousands, fill = AgeGroup))+
  geom_area(position = "fill", color = "black", alpha = .5)+
  scale_fill_manual(values = brewer.pal(8, "Greens"))
```

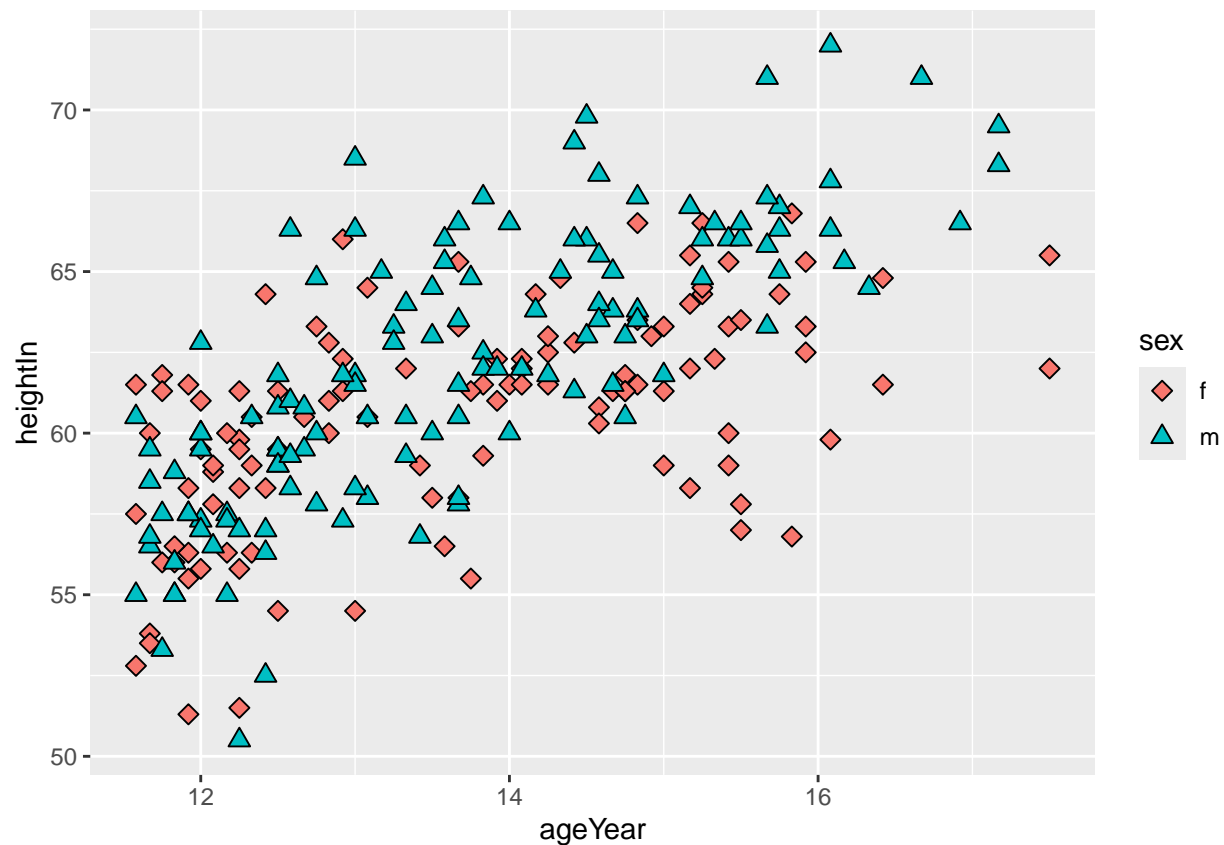


Nube de puntos

```
head(heightweight)
```

```
##   sex ageYear ageMonth heightIn weightLb
## 1  f   11.92    143     56.3     85.0
## 2  f   12.92    155     62.3    105.0
## 3  f   12.75    153     63.3    108.0
## 4  f   13.42    161     59.0     92.0
## 5  f   15.92    191     62.5    112.5
## 6  f   14.25    171     62.5    112.0
```

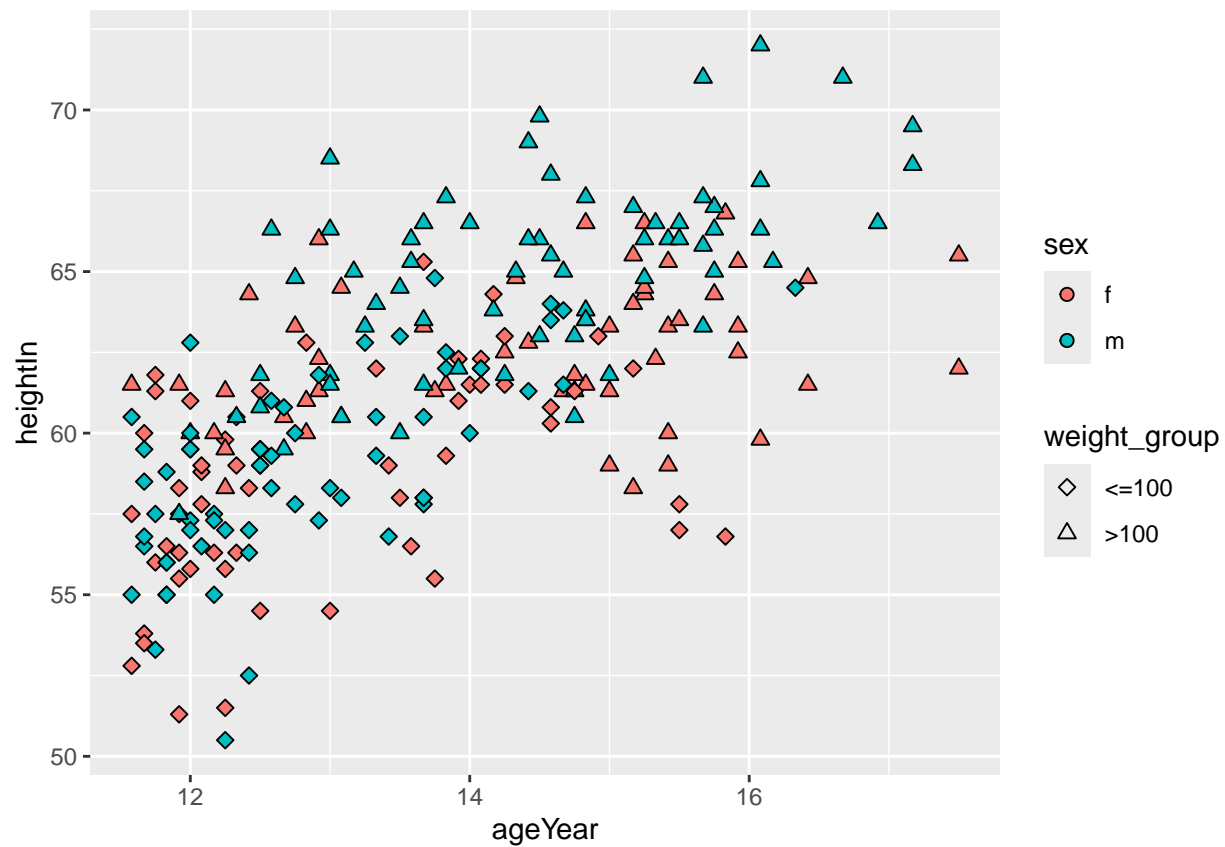
```
ggplot(heightweight, aes(x=ageYear, y=heightIn, shape = sex, fill= sex)) +
  geom_point(size=2.5)+
  scale_shape_manual(values = c(23,24))
```



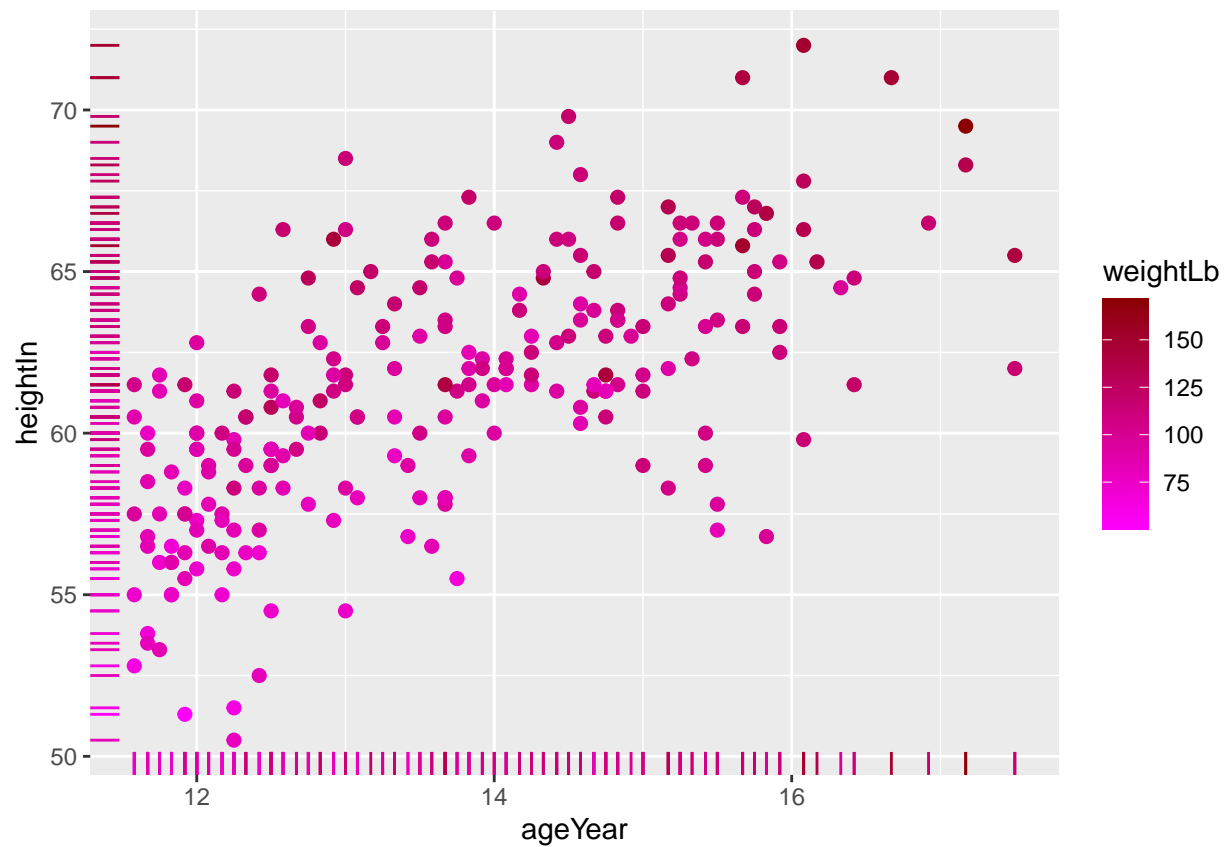
```
heightweight$weight_group = ifelse(heightweight$weightLb <= 100, "<=100", ">100")
head(heightweight)
```

```
##   sex ageYear ageMonth heightIn weightLb weight_group
## 1  f   11.92    143    56.3     85.0      <=100
## 2  f   12.92    155    62.3    105.0      >100
## 3  f   12.75    153    63.3    108.0      >100
## 4  f   13.42    161    59.0     92.0      <=100
## 5  f   15.92    191    62.5    112.5      >100
## 6  f   14.25    171    62.5    112.0      >100
```

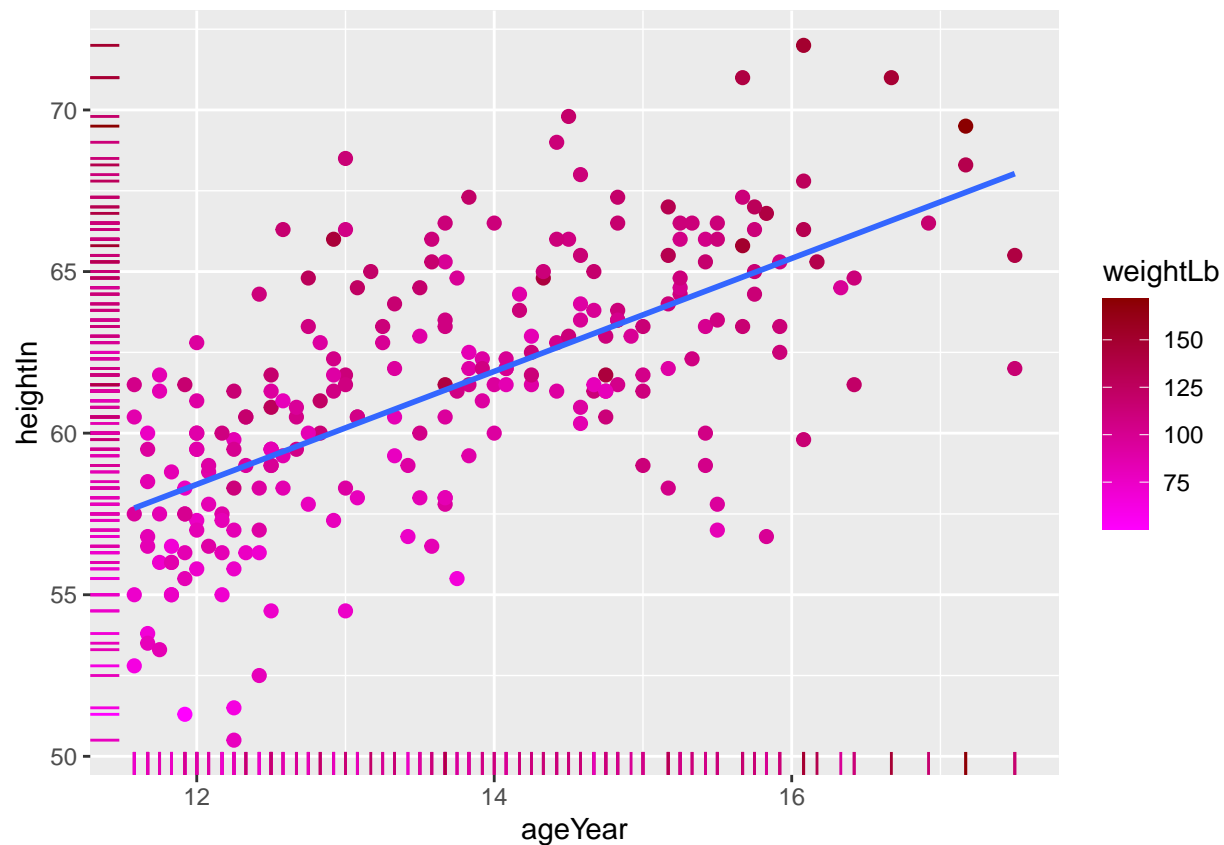
```
ggplot(heightweight, aes(x=ageYear, y=heightIn, shape = weight_group, fill = sex, color=sex))+
  geom_point(size=2, color = "black")+
  scale_shape_manual(values = c(23,24))+
  guides(fill = guide_legend(override.aes = c(shape=21)))
```



```
ggplot(heightweight, aes(x=ageYear, y=heightIn, color = weightLb))+
  geom_point(size=2)+
  geom_rug()+
  scale_color_gradient(low = "magenta", high = "red4")
```



```
ggplot(heightweight, aes(x=ageYear, y=heightIn, color = weightLb))+
  geom_point(size=2)+
  geom_rug()+
  scale_color_gradient(low = "magenta", high = "red4")+
  geom_smooth(method = "lm", se=FALSE)
```

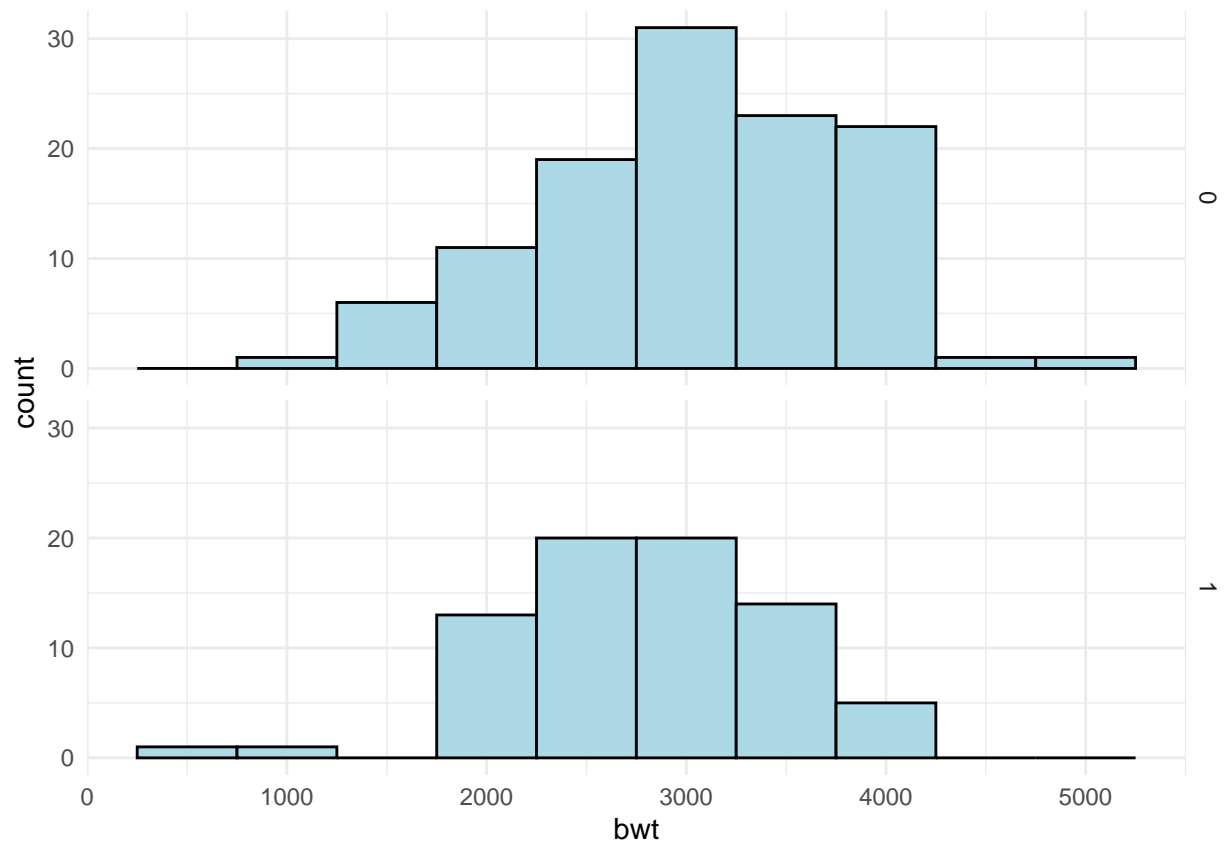


Histograma

```
head(birthwt)
```

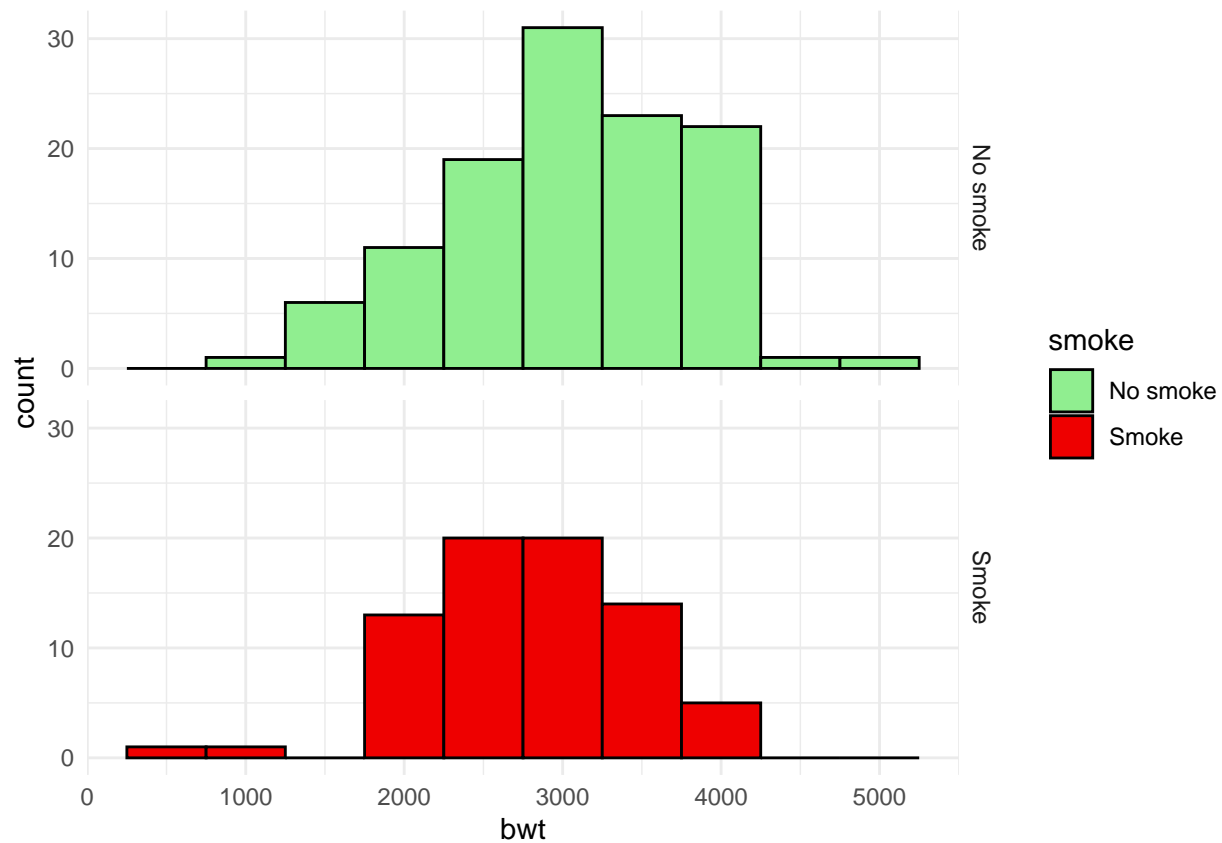
```
##      low age lwt race smoke ptl ht ui ftv  bwt
## 85    0  19 182    2     0   0  0  1   0 2523
## 86    0  33 155    3     0   0  0  0   3 2551
## 87    0  20 105    1     1   0  0  0   1 2557
## 88    0  21 108    1     1   0  0  1   2 2594
## 89    0  18 107    1     1   0  0  1   0 2600
## 91    0  21 124    3     0   0  0  0   0 2622
```

```
ggplot(birthwt, aes(x=bwt))+
  geom_histogram(fill = "lightblue", color = "black", binwidth = 500 )+
  facet_grid(smoke~.)+
  theme_minimal()
```

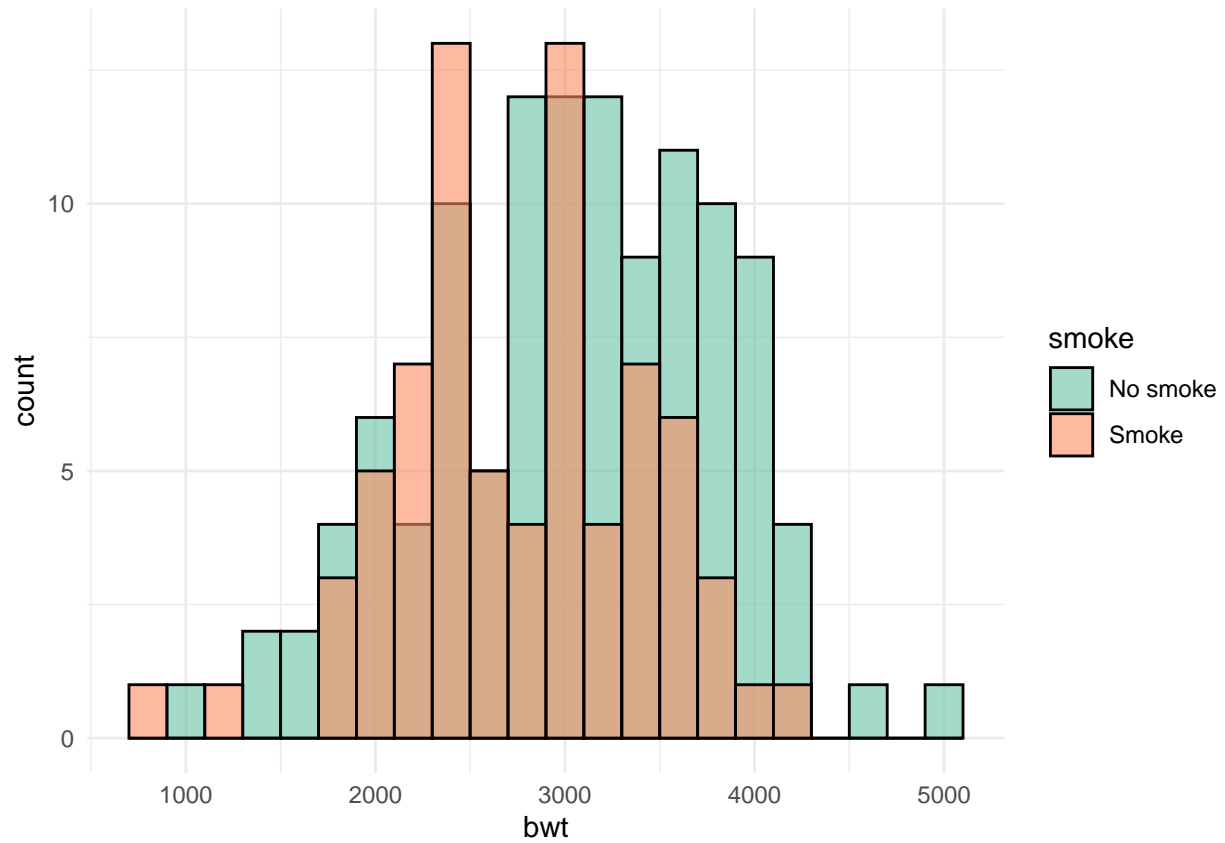


```
birthwt2 = birthwt
birthwt2$smoke = factor(ifelse(birthwt2$smoke == 0, "No smoke", "Smoke"))

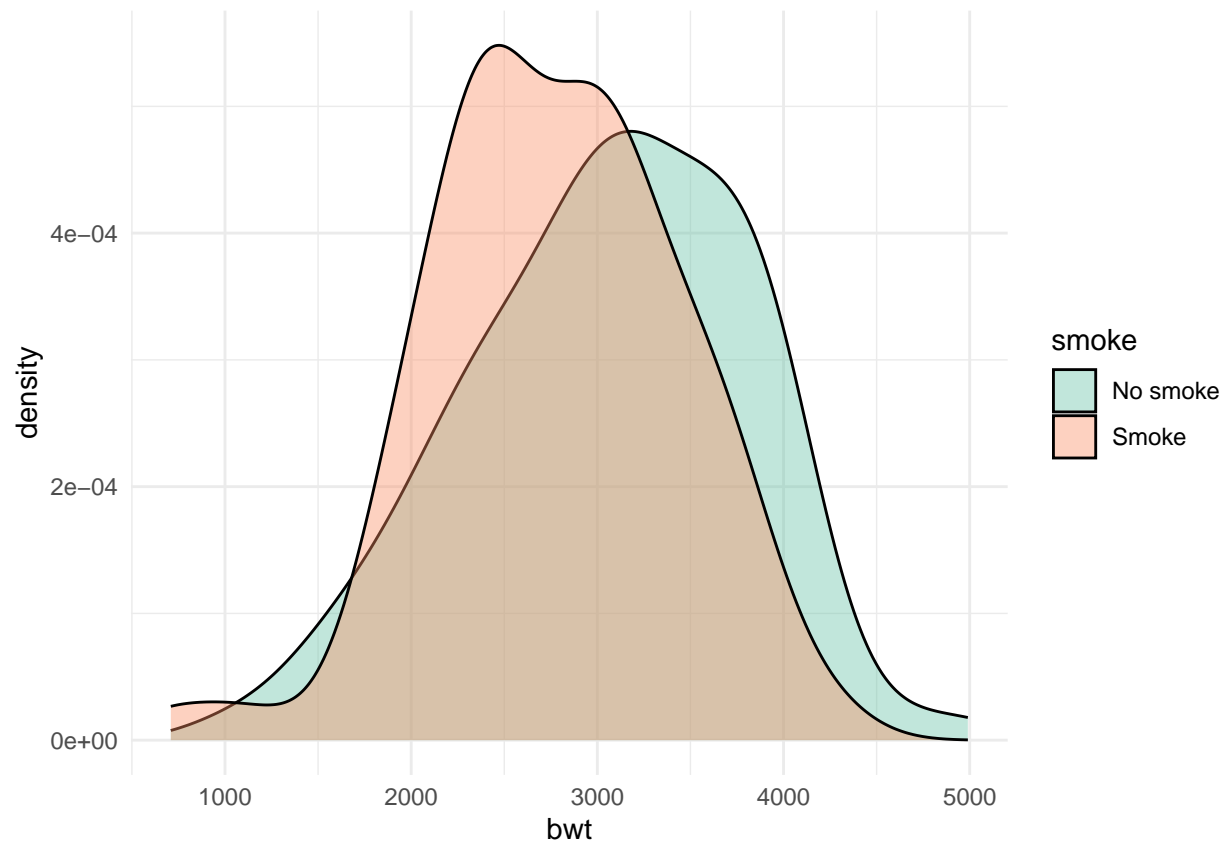
ggplot(birthwt2, aes(x=bwt, fill = smoke))+
  geom_histogram(color = "black", binwidth = 500 )+
  facet_grid(smoke~.)+
  scale_fill_manual(values = c("lightgreen", "red2"))+
  theme_minimal()
```

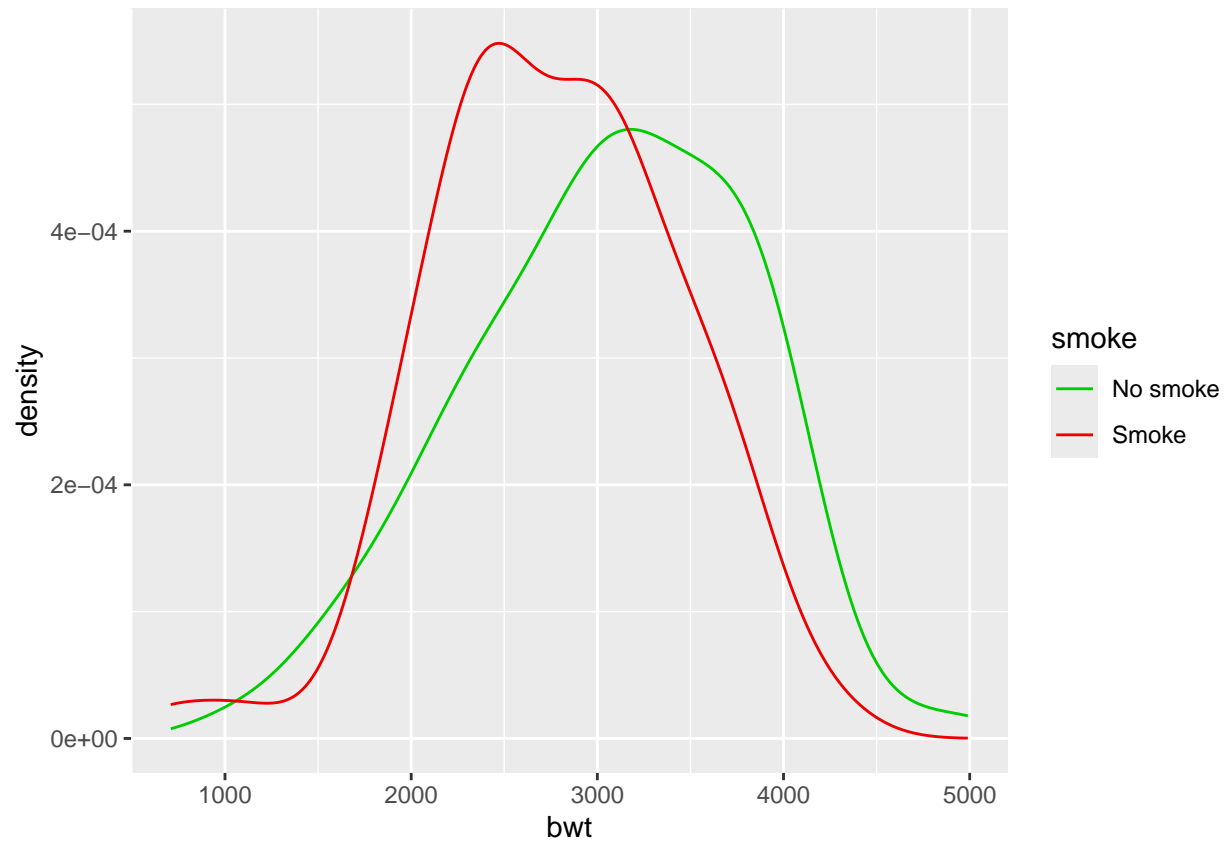
```
ggplot(birthwt2, aes(x=bwt, fill = smoke))+  
  geom_histogram(color = "black", binwidth = 200, position = "identity", alpha = 0.6)+  
  scale_fill_manual(values = brewer.pal(n=2, name="Set2"))+  
  theme_minimal()
```



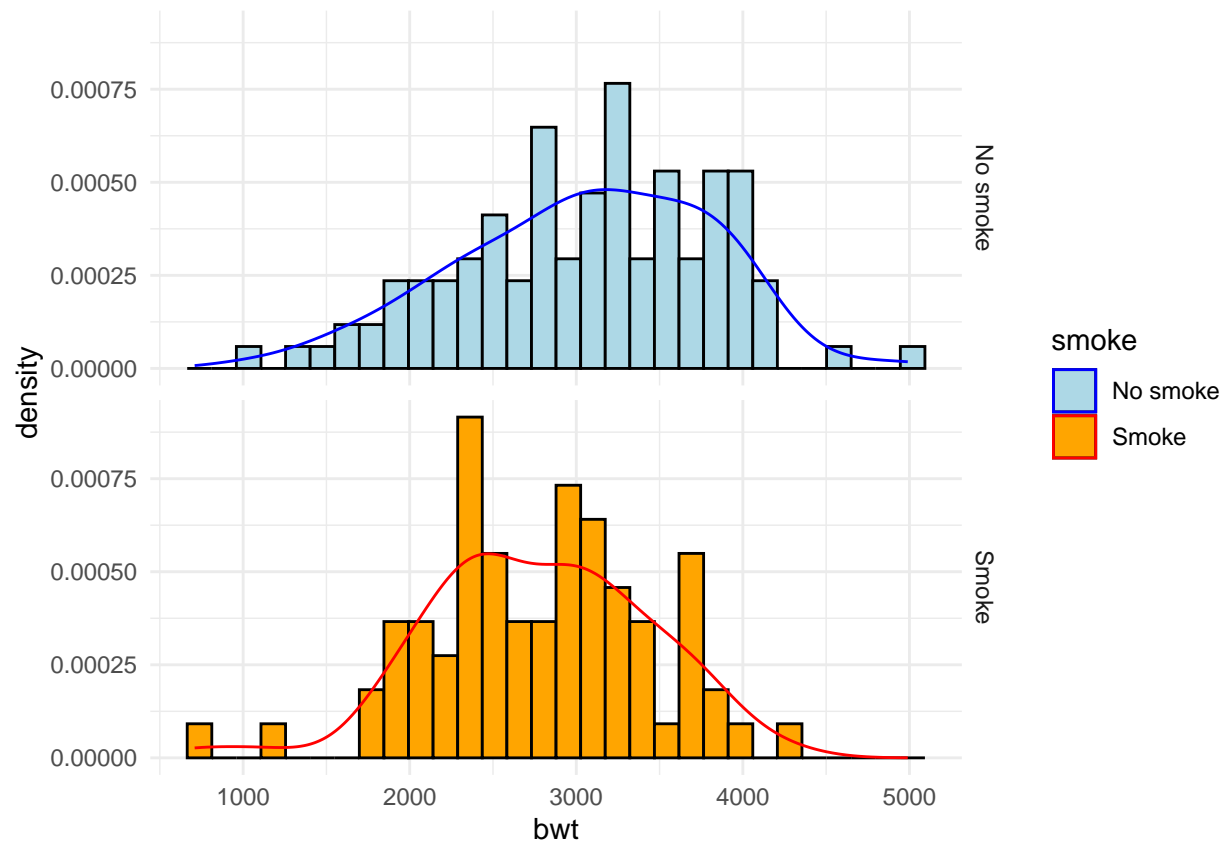
```
ggplot(birthwt2, aes(x=bwt, fill = smoke))+  
  geom_density(alpha = 0.4)+  
  scale_fill_manual(values = brewer.pal(n=2, name="Set2"))+  
  theme_minimal()
```



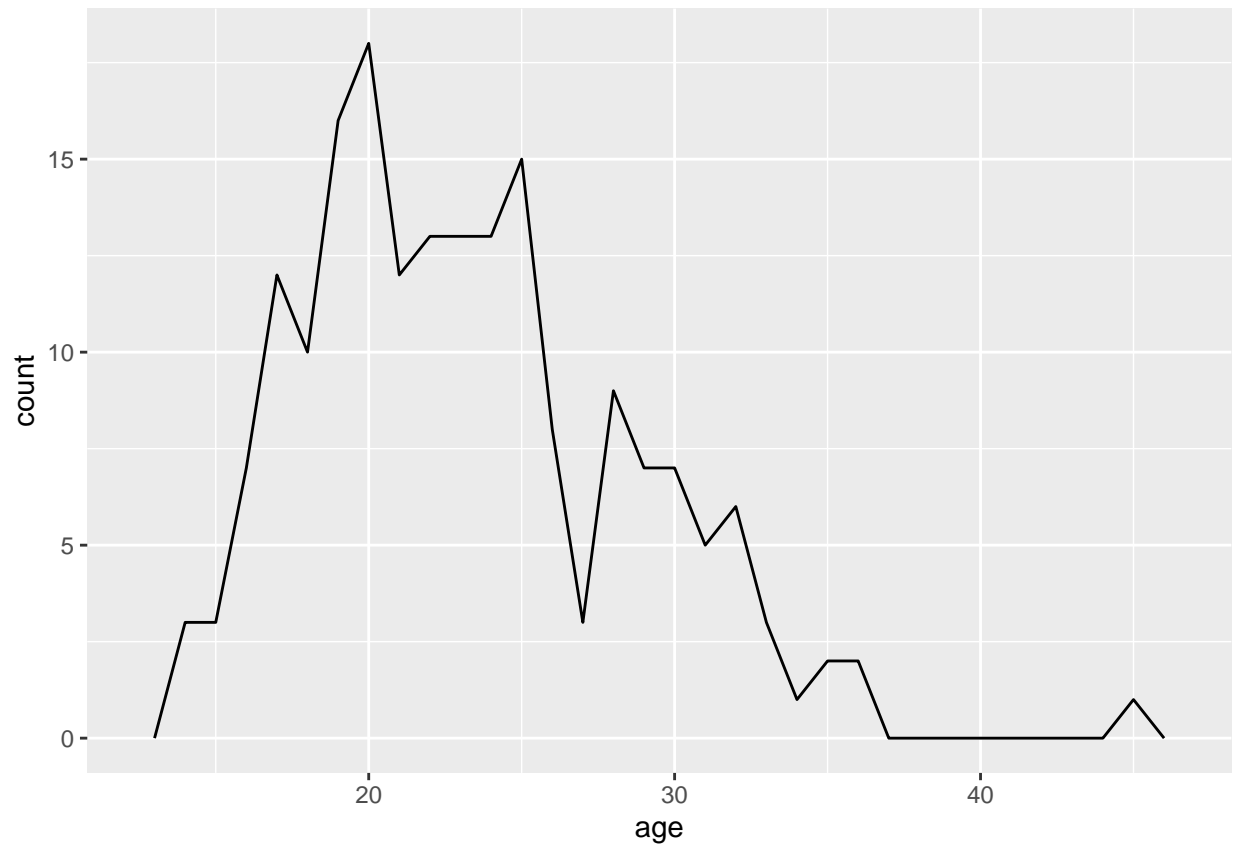
```
ggplot(birthwt2, aes(x=bwt, color = smoke))+  
  geom_line(stat = "density")+  
  scale_color_manual(values = c("green3", "red2"))
```



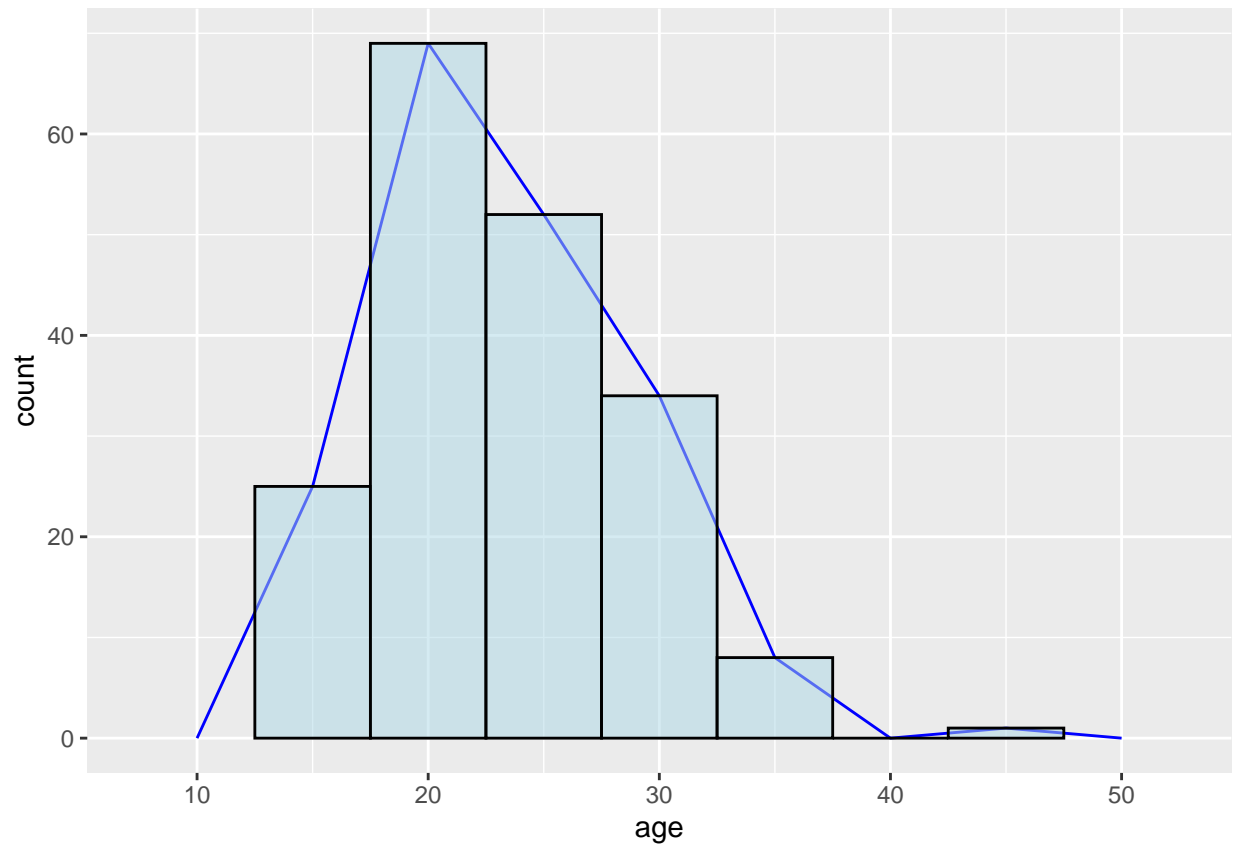
```
ggplot(birthwt2, aes(x=bwt, fill = smoke, color = smoke))+
  geom_histogram(aes(y = ..density..), color = "black")+
  geom_density(fill = "white", alpha = 0)+
  facet_grid(smoke~.)+
  scale_fill_manual(values = c("lightblue", "orange"))+
  scale_color_manual(values = c("blue", "red"))+
  theme_minimal()
```



```
ggplot(birthwt2, aes(x=bwt)) +  
  geom_freqpoly(binwidth = 1)
```

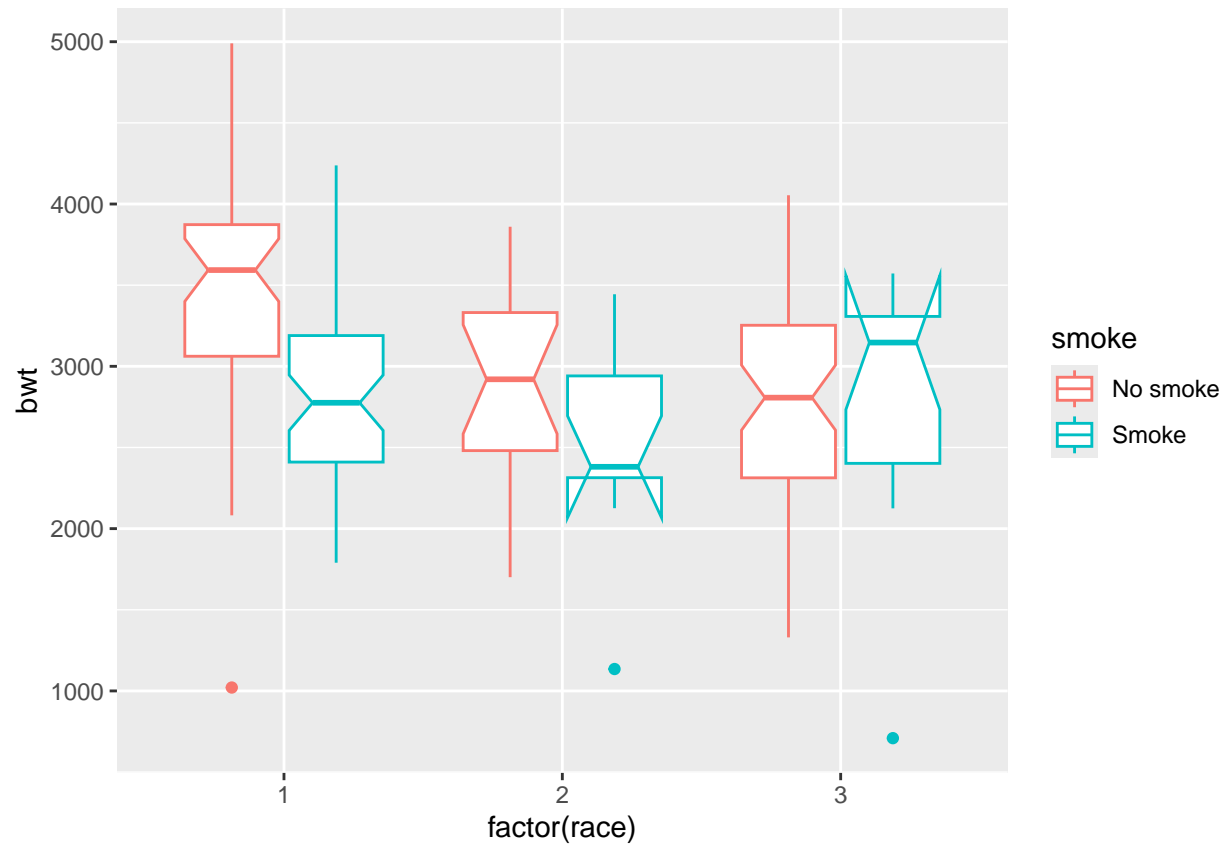


```
ggplot(birthwt2, aes(x=age))+  
  geom_freqpoly(binwidth = 5, color = "blue")+  
  geom_histogram(binwidth = 5, alpha = 0.5, fill = "lightblue", color = "black")
```

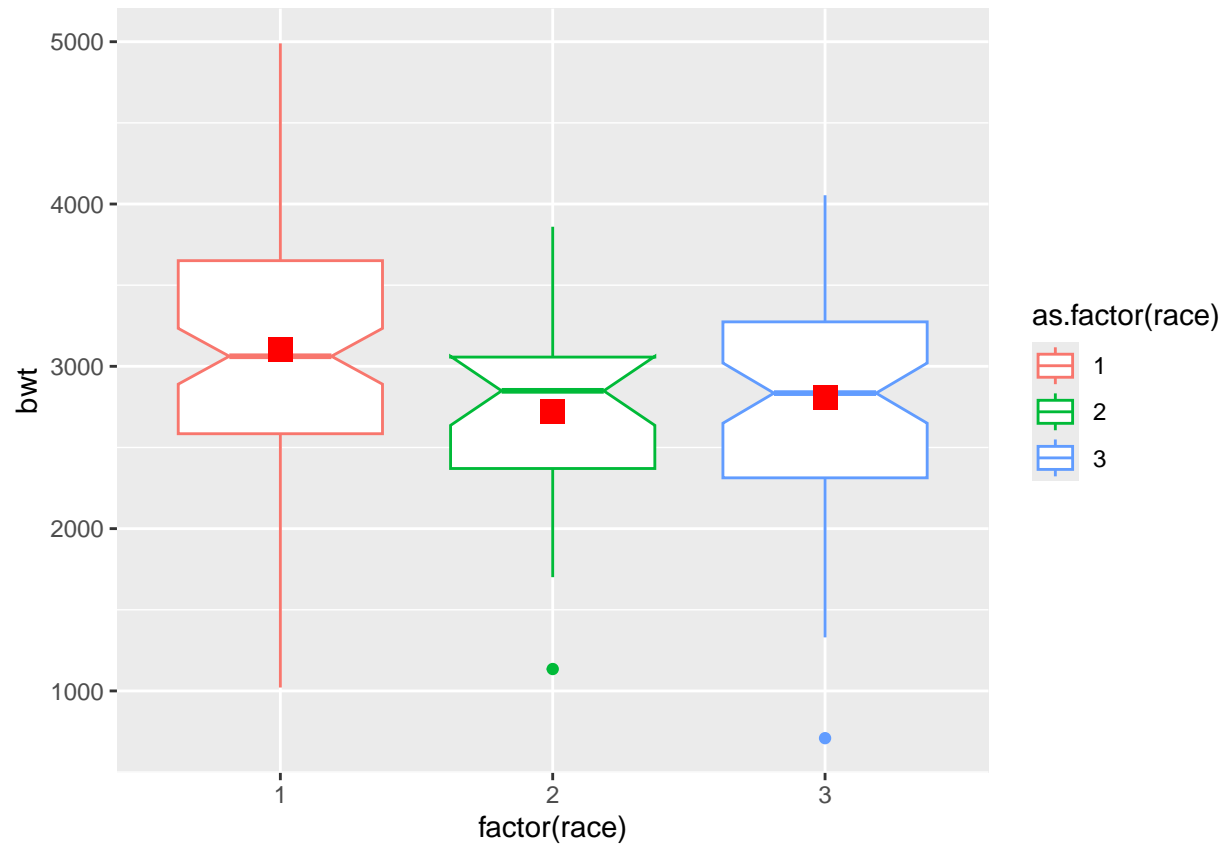


boxplot

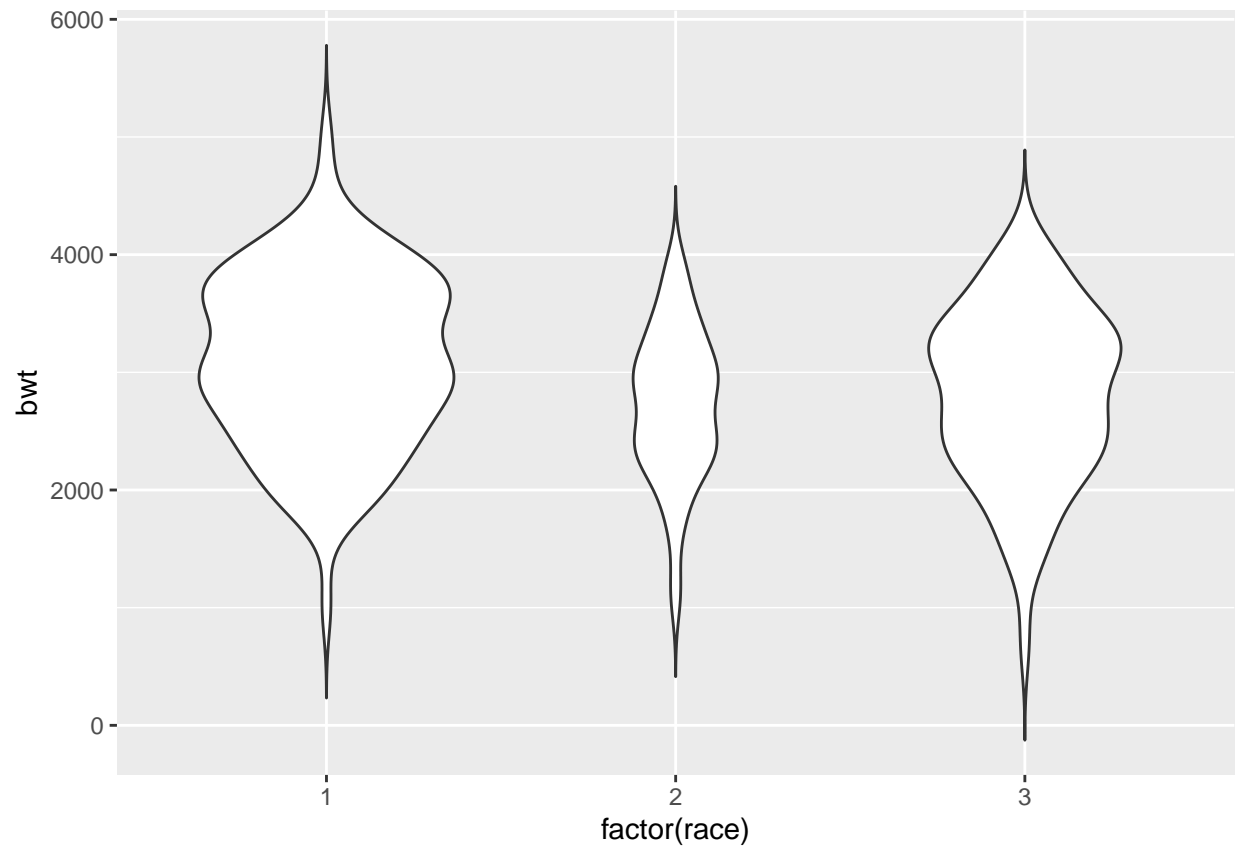
```
ggplot(birthwt2, aes(x=factor(race), y = bwt, color = smoke))+  
  geom_boxplot(notch = TRUE)
```



```
ggplot(birthwt2, aes(x=factor(race), y = bwt, color = as.factor(race)))+  
  geom_boxplot(notch = TRUE)+  
  stat_summary(fun.y=mean, geom="point", shape=15, size=4, color="red")
```

```
ggplot(birthwt2, aes(x = factor(race), y = bwt)) +  
  geom_violin(trim = FALSE, scale = "count")
```

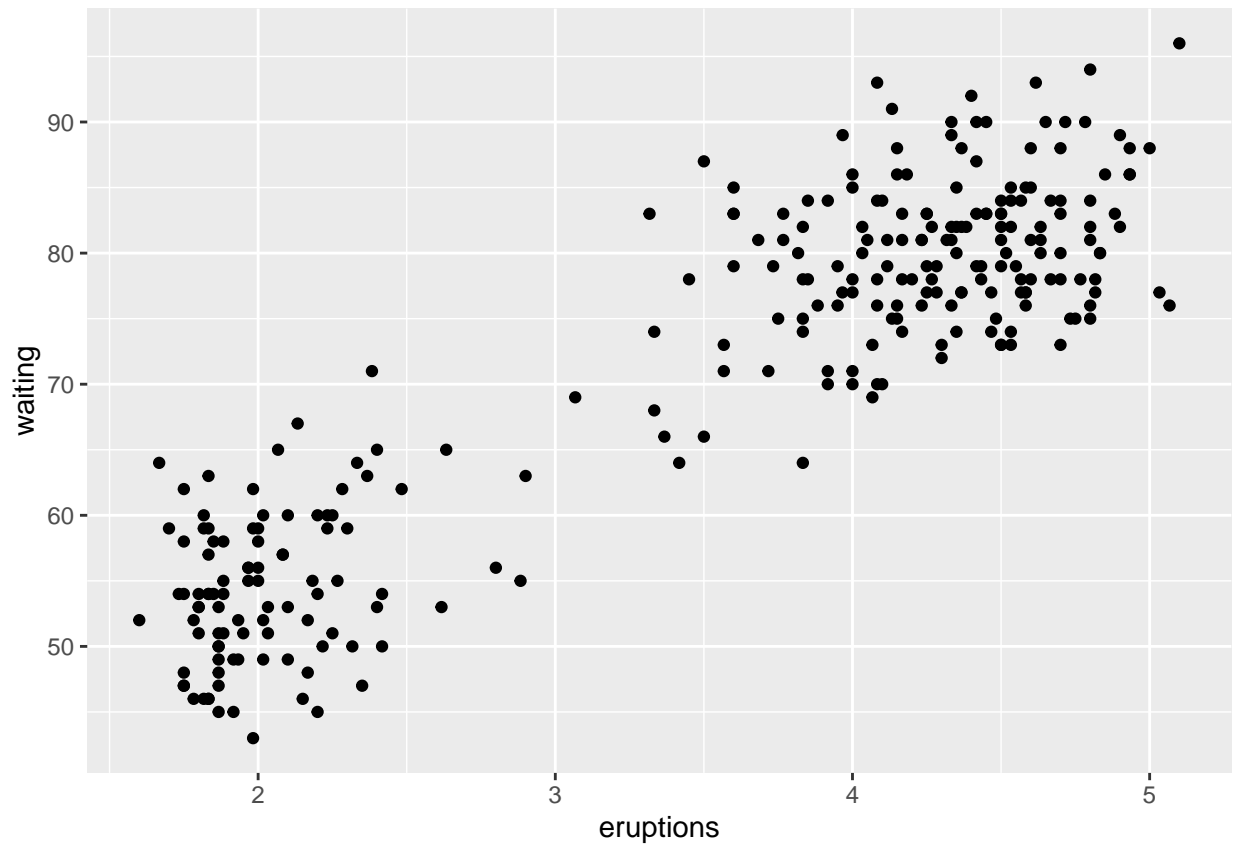


Density 2D

```
head(faithful)
```

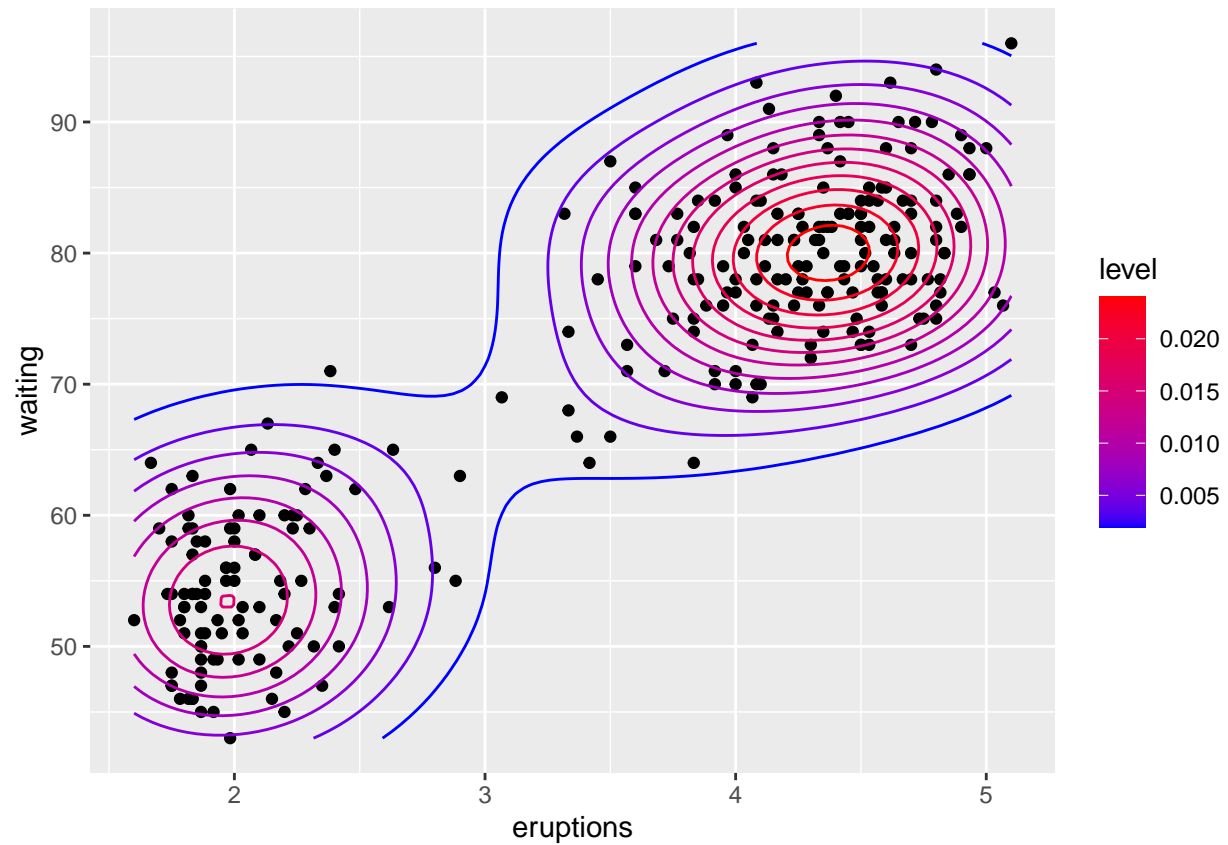
```
##   eruptions waiting
## 1     3.600      79
## 2     1.800      54
## 3     3.333      74
## 4     2.283      62
## 5     4.533      85
## 6     2.883      55
```

```
ggplot(faithful, aes(x= eruptions, y=waiting))+
  geom_point()
```

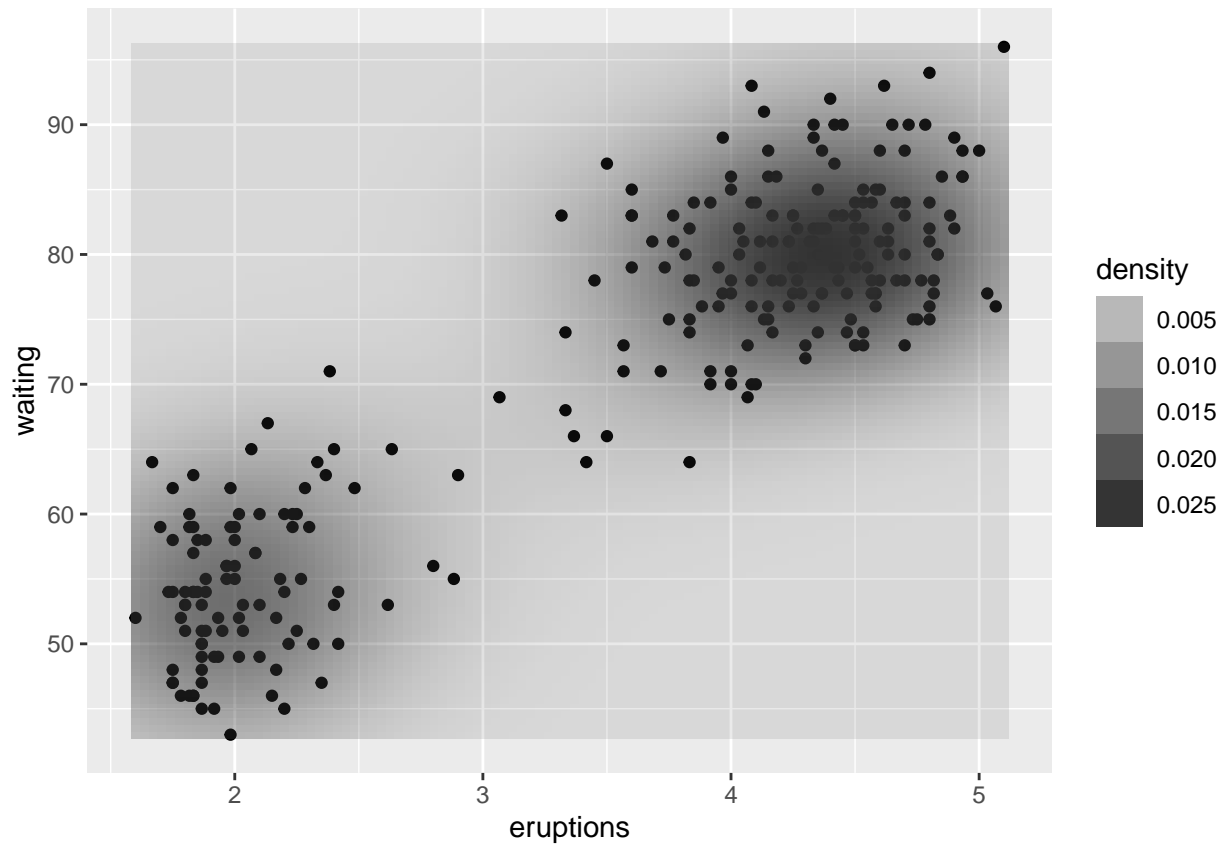


```
# stat_density_2d()
```

```
ggplot(faithful, aes(x= eruptions, y=waiting))+  
  geom_point()+  
  stat_density_2d(aes(color=..level..))+  
  scale_color_gradient(low = "blue", high = "red")
```

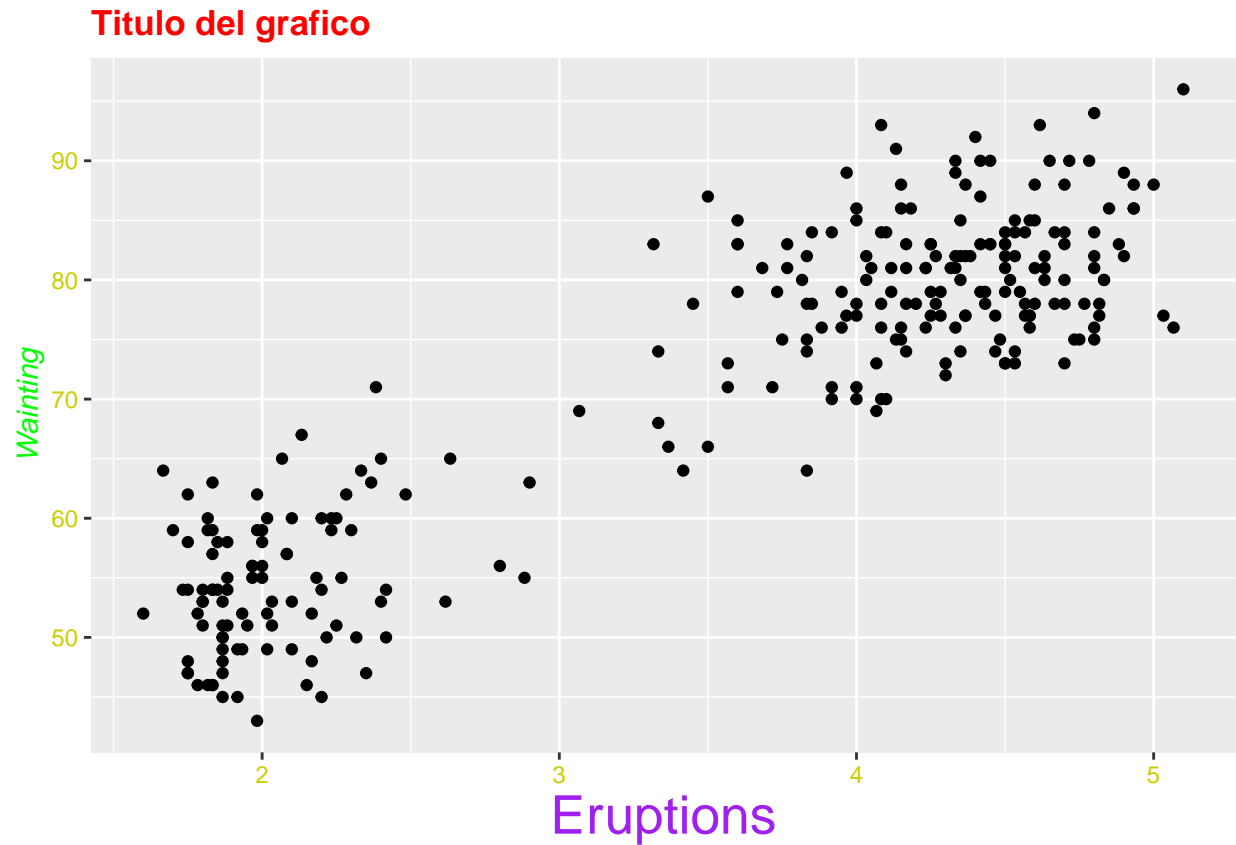


```
ggplot(faithful, aes(x= eruptions, y=waiting))+  
  geom_point()+  
  stat_density_2d(aes(alpha=..density..), geom = 'raster', contour = F)
```

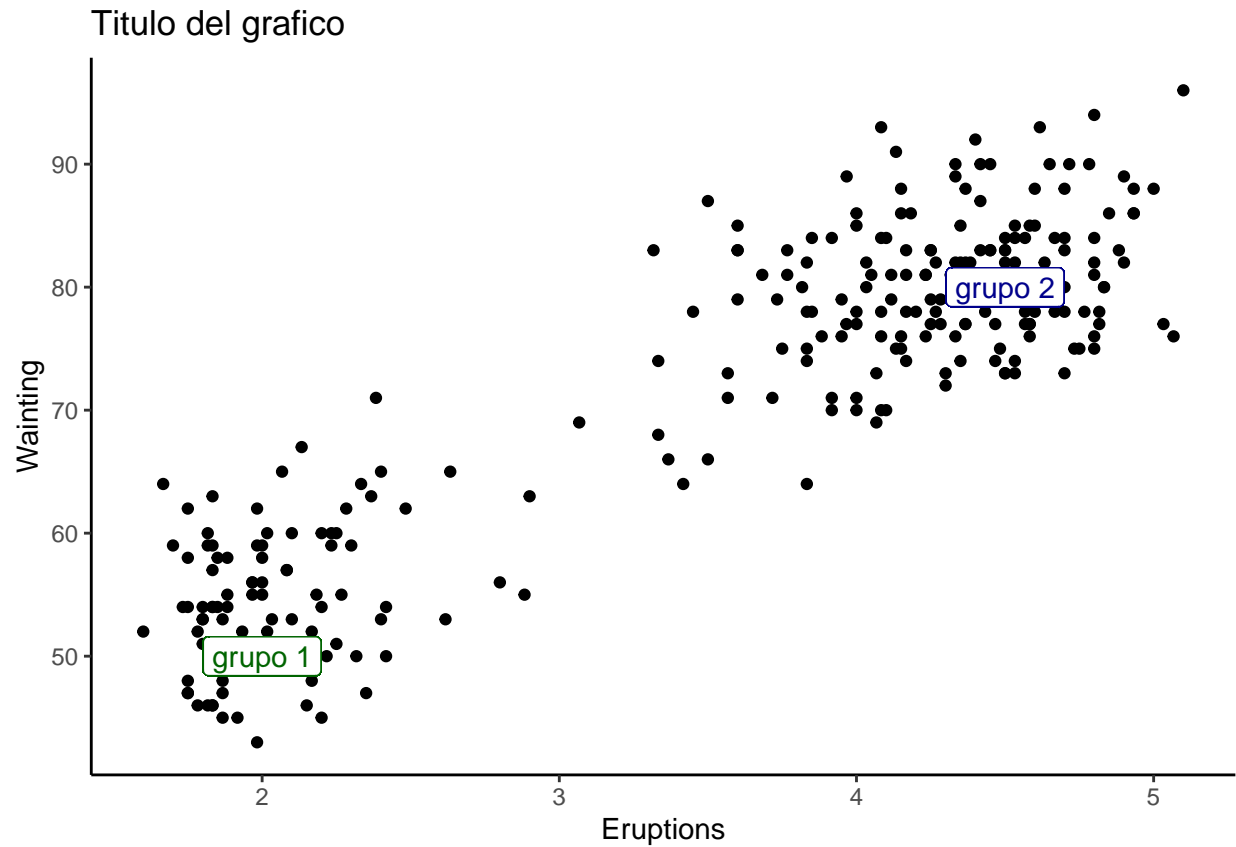


Modifier les éléments d'un graphique

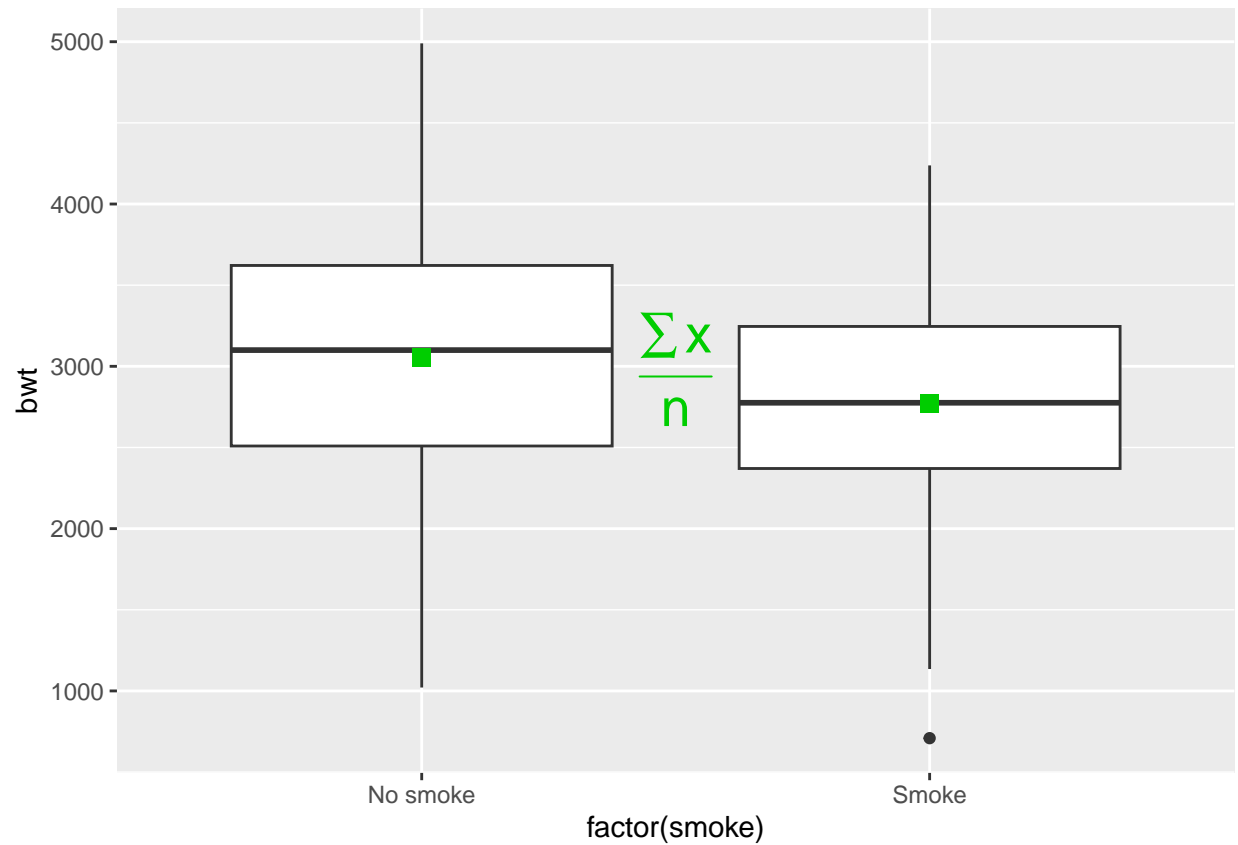
```
ggplot(faithful, aes(x= eruptions, y=waiting))+
  geom_point()+
  xlab("Eruptions") +
  ylab("Waiting")+
  ggtitle("Titulo del grafico")+
  theme(plot.title = element_text(
    color = "red",
    face = "bold"
  ),
  axis.title.x = element_text(
    color = "purple",
    size = 20
  ),
  axis.title.y = element_text(
    color = "green",
    face = "italic"
  ),
  axis.text = element_text(
    color = "yellow3"
  ))
```



```
ggplot(faithful, aes(x= eruptions, y=waiting))+  
  geom_point()+  
  xlab("Eruptions") +  
  ylab("Waiting")+  
  ggtitle("Titulo del grafico")+  
  geom_label(label = "grupo 1",  
             x = 2,  
             y = 50,  
             color = "darkgreen")+  
  geom_label(label = "grupo 2",  
             x = 4.5,  
             y = 80,  
             color = "darkblue")+  
  theme_classic()
```

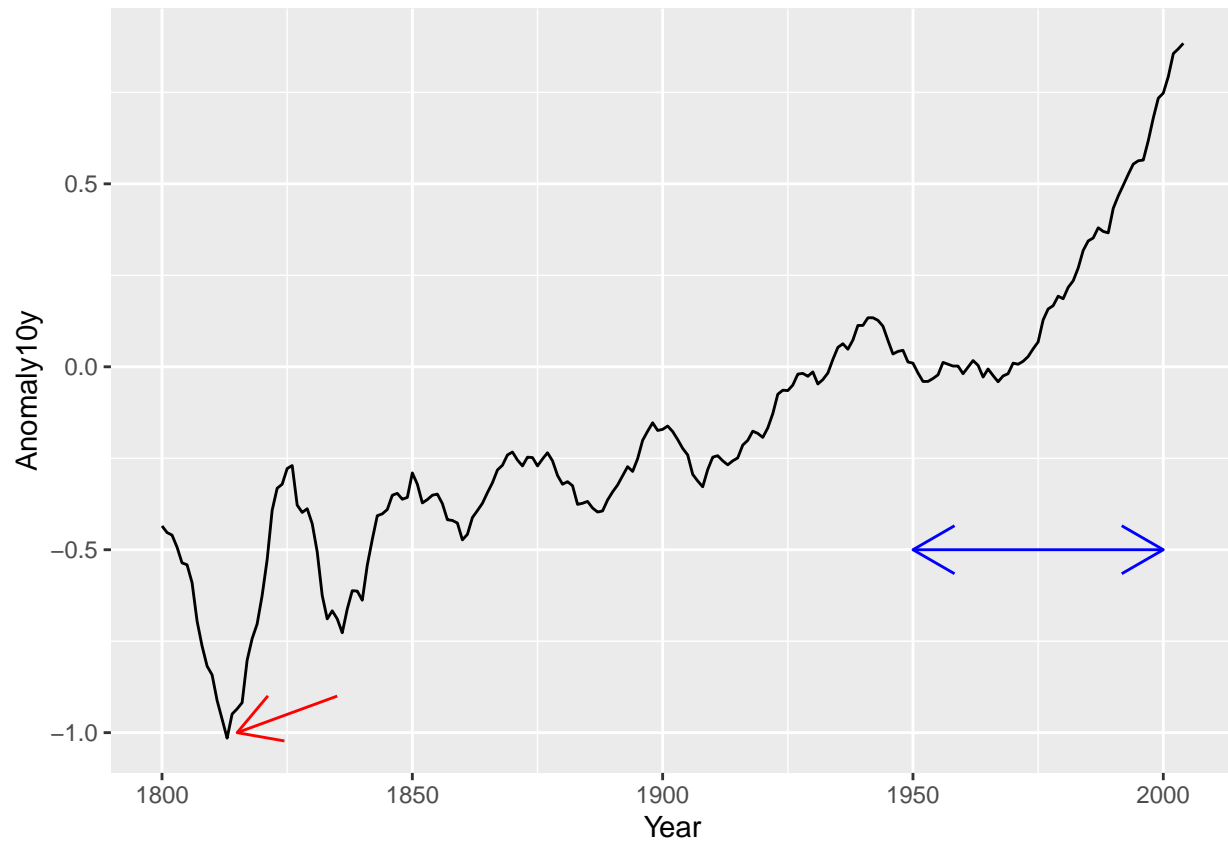


```
ggplot(birthwt2, aes(x=factor(smoke), y=bwt))+
  geom_boxplot()+
  stat_summary(fun = "mean",
               geom = "point",
               size = 3,
               color = "green3",
               shape = 15)+
  annotate(geom = "text",
           x = 1.5,
           y = 3000,
           size = 7,
           label = expression(frac(sum(x),n)),
           color = "green3")
```



```
datos = climate %>%
  filter(Source == "Berkeley")

ggplot(datos, aes(x=Year, y = Anomaly10y))+
  geom_line()+
  annotate("Segment",
    x = 1835, y = -0.9,
    xend = 1815, yend = -1,
    arrow = arrow(),
    color = "red")+
  annotate("segment",
    x = 1950, xend = 2000,
    y = -0.5, yend = -0.5,
    arrow = arrow(ends = "both"),
    color = "blue")
```

WordCloud

```
head(demoFreq)
```

```
##      word freq
## oil      oil   85
## said     said  73
## prices  prices 48
## opec     opec  42
## mln      mln   31
## the      the   26
```

```
# wordcloud2(data = demoFreq, size = 1.6)
```

```
# wordcloud2(data = demoFreq, size = 1.6,
#            color = paletteer_d("tvthemes::Day"),
#            backgroundColor = "lightyellow")
```

```
# wordcloud2(data = demoFreq,
#            figPath = "apple.png")
```

The shape of the "cloud" to draw. Can be a keyword present. Available presents are 'circle' (default)