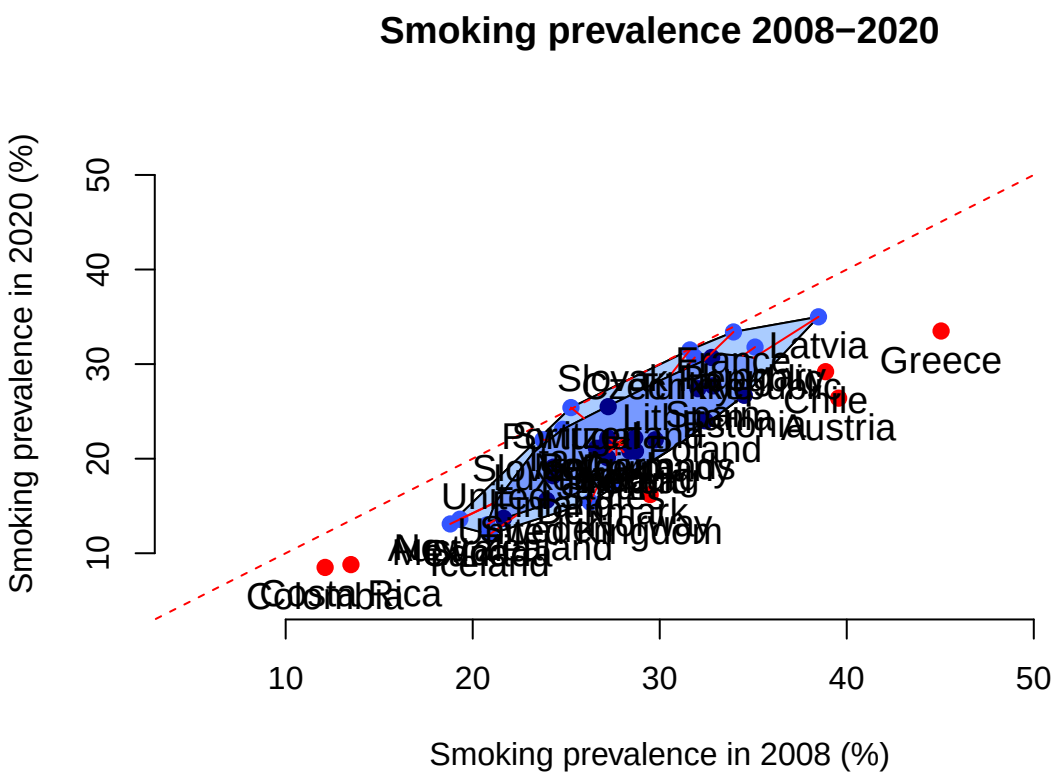
# Prevalence analysis

2023-07-05

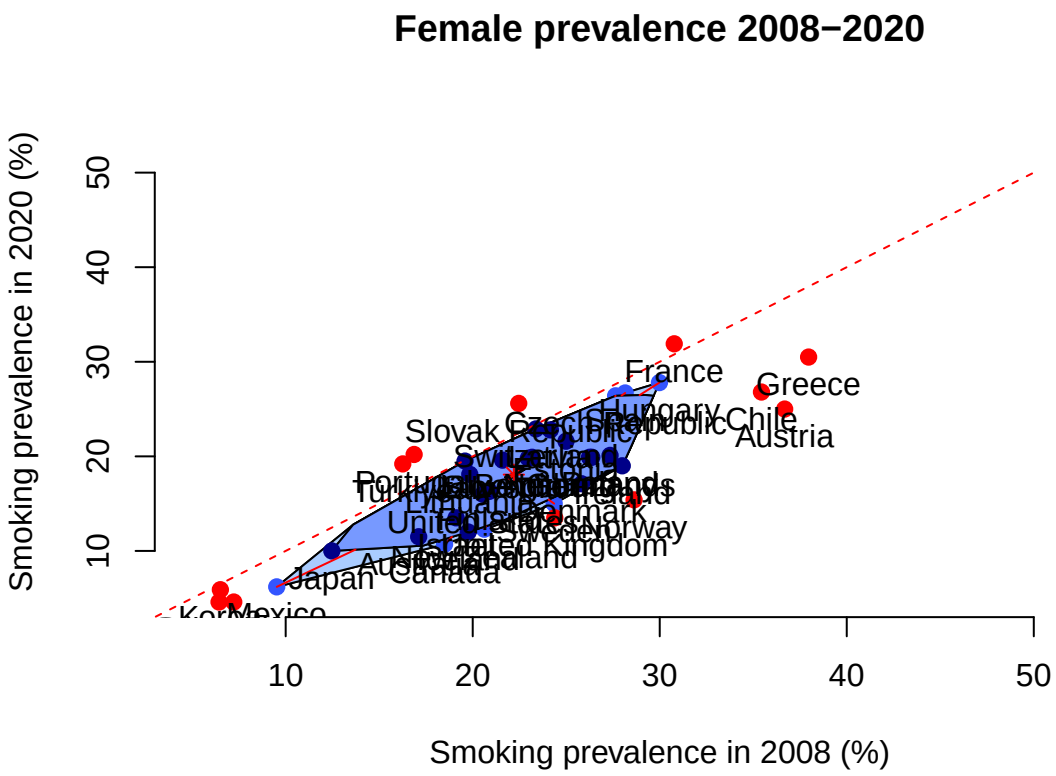
We plot the bagplots for overall, male and female prevalence

```
#Both
bagplot(df1,
        factor=1.5,
        xlim=c(5,55),
        ylim=c(5,55),
        cex=1.2,
        xlab='Smoking prevalence in 2008 (%)',
        ylab="Smoking prevalence in 2020 (%)",
        main="Smoking prevalence 2008-2020")
par(cex.axis = 1.4,cex.lab=1.5,cex.main=1.8)
lines(x=0:50, y=0:50, col='red',lty=2)
text(data_2008,
     data_2020,
     pos=1,
     cex=1.2,
     labels = countries)
```




Females

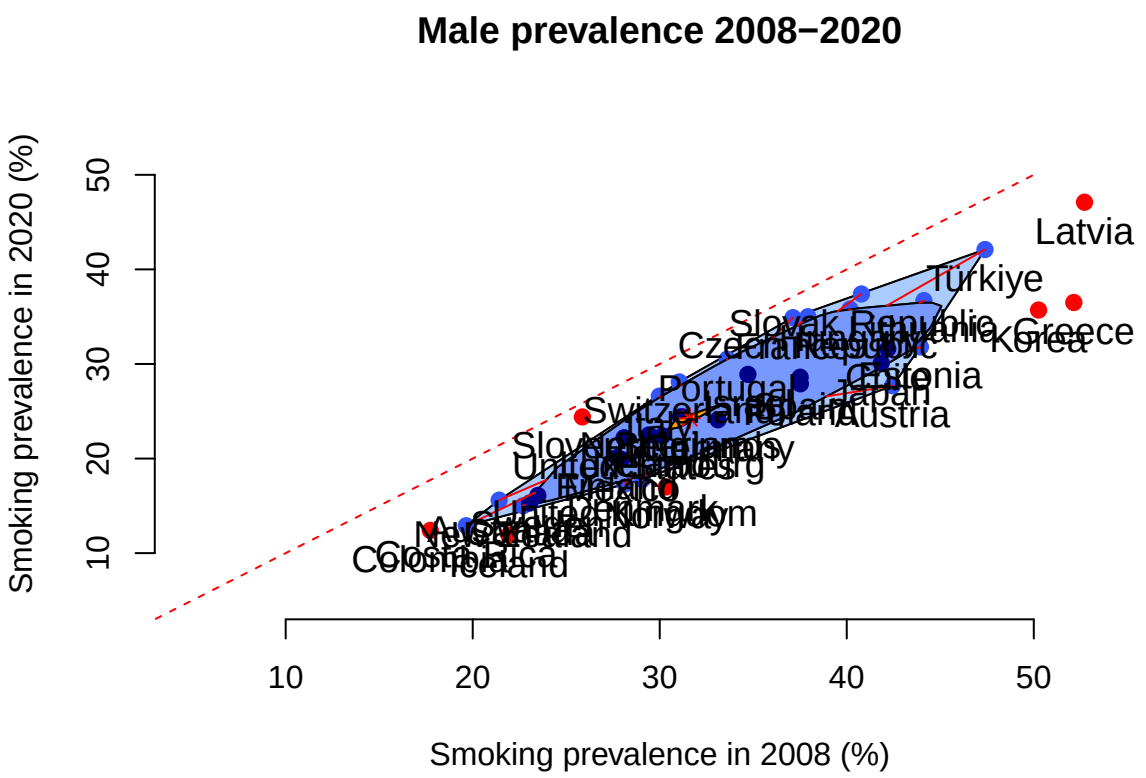
```
#Females
bagplot(df_f,
        factor=1.5,
        xlim=c(5,55),
        ylim=c(5,55),
        cex=1.2,
        xlab='Smoking prevalence in 2008 (%)',
        ylab="Smoking prevalence in 2020 (%)",
        main="Female prevalence 2008-2020")
par(cex.axis = 1.4,cex.lab=1.5,cex.main=1.8)
lines(x=0:50, y=0:50, col='red',lty=2)
text(data_2008_females,
     data_2020_females,
     pos=1,
     labels = countries)
```




Males

```
#Males
bagplot(df_m,
        factor=1.5,
        xlim=c(5,55),
        ylim=c(5,55),
        cex=1.2,
        xlab='Smoking prevalence in 2008 (%)',
        ylab="Smoking prevalence in 2020 (%)",
        main="Male prevalence 2008-2020")
par(cex.axis = 1.4,cex.lab=1.5,cex.main=1.8)
lines(x=0:50, y=0:50, col='red',lty=2)
```

```
text(data_2008_males,
     data_2020_males,
     pos=1,
     labels = countries,
     cex=1.2)
```

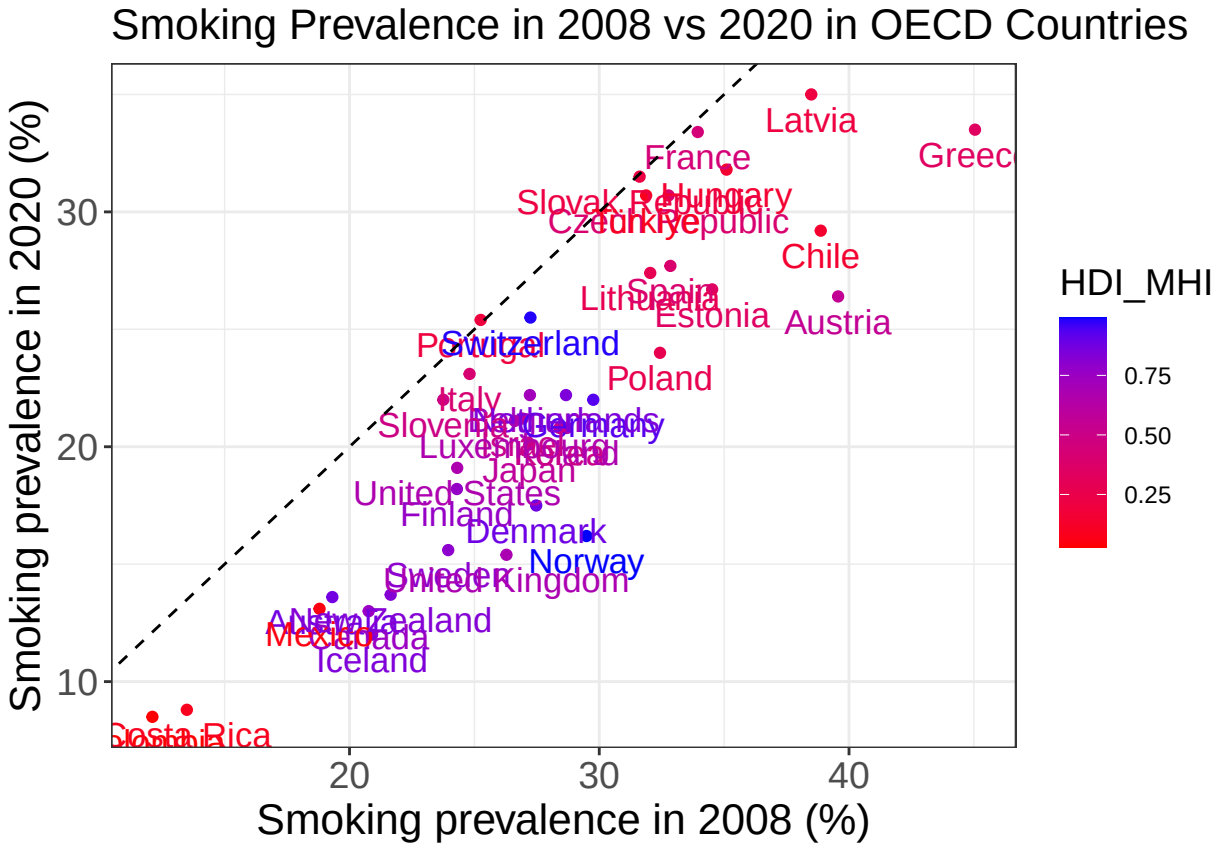



Now we plot the scatterplot using the HDI index. We start from the overall prevalence

```
databub<-data.frame(data_2008=data_2008,
                    data_2020=data_2020,
                    HDI_MHI=HDI_MHI)


# Create a 2D bubble plot for both sexes
ggplot(data=databub,aes(x = data_2008, y = data_2020,  color = HDI_MHI)) +
  geom_point() +
  geom_text(aes(label = countries), vjust = 1.5, size=4.5) +  # Add labels with the 'Country' variable
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "black") +  # Add the y = x line
  scale_color_gradient(low = "red", high = "blue") +
  labs(x = "Smoking prevalence in 2008 (%)",
       y = "Smoking prevalence in 2020 (%)",
       size = "Population",
       color = "HDI_MHI") +
  theme_bw() +
  theme(
    axis.title = element_text(size = 16),
    axis.text = element_text(size = 14),  # Increase the size of the axis labels
    plot.title = element_text(size = 16),
    legend.title = element_text(size = 14),
```

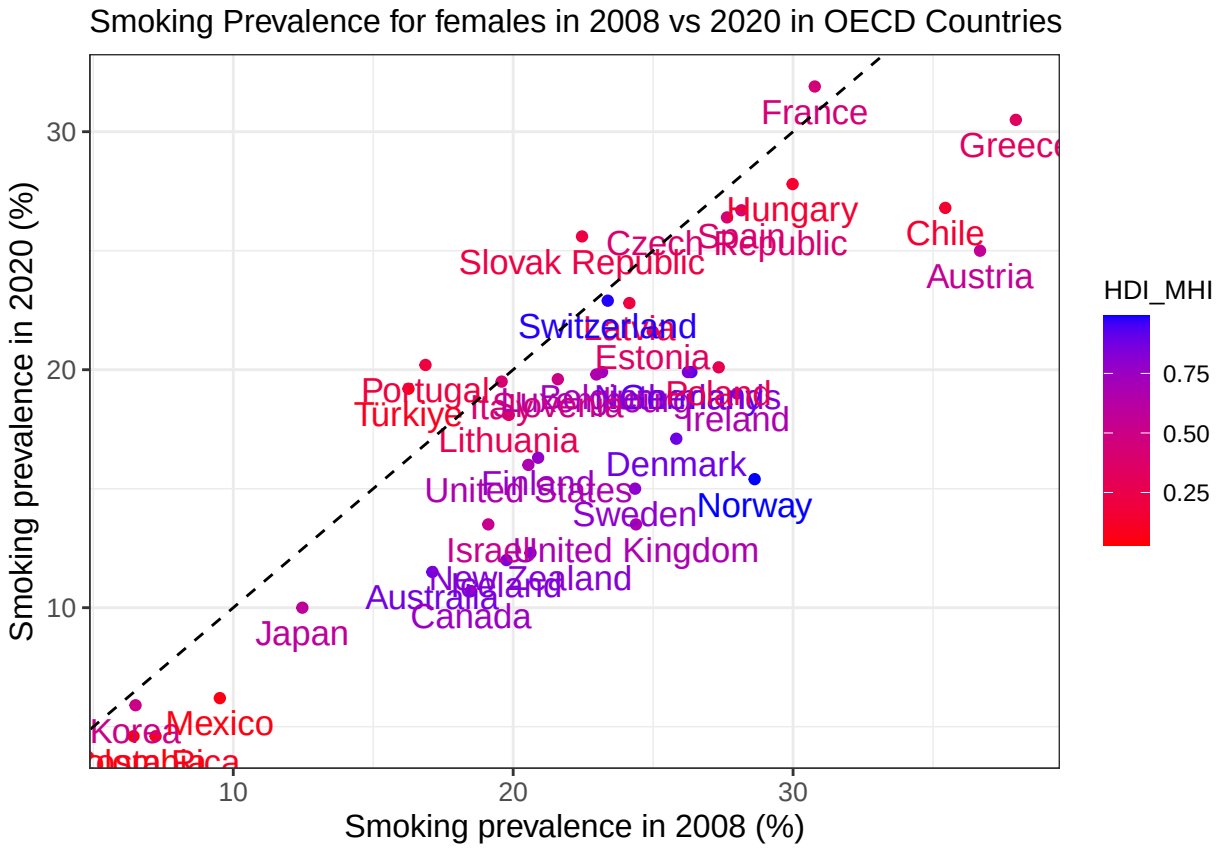
```
    panel.background = element_rect(fill = "white")  # Set a white color for the panel background
  ) +
  # Increase the size of the color label
  ggtitle("Smoking Prevalence in 2008 vs 2020 in OECD Countries")
```




We now analyze females

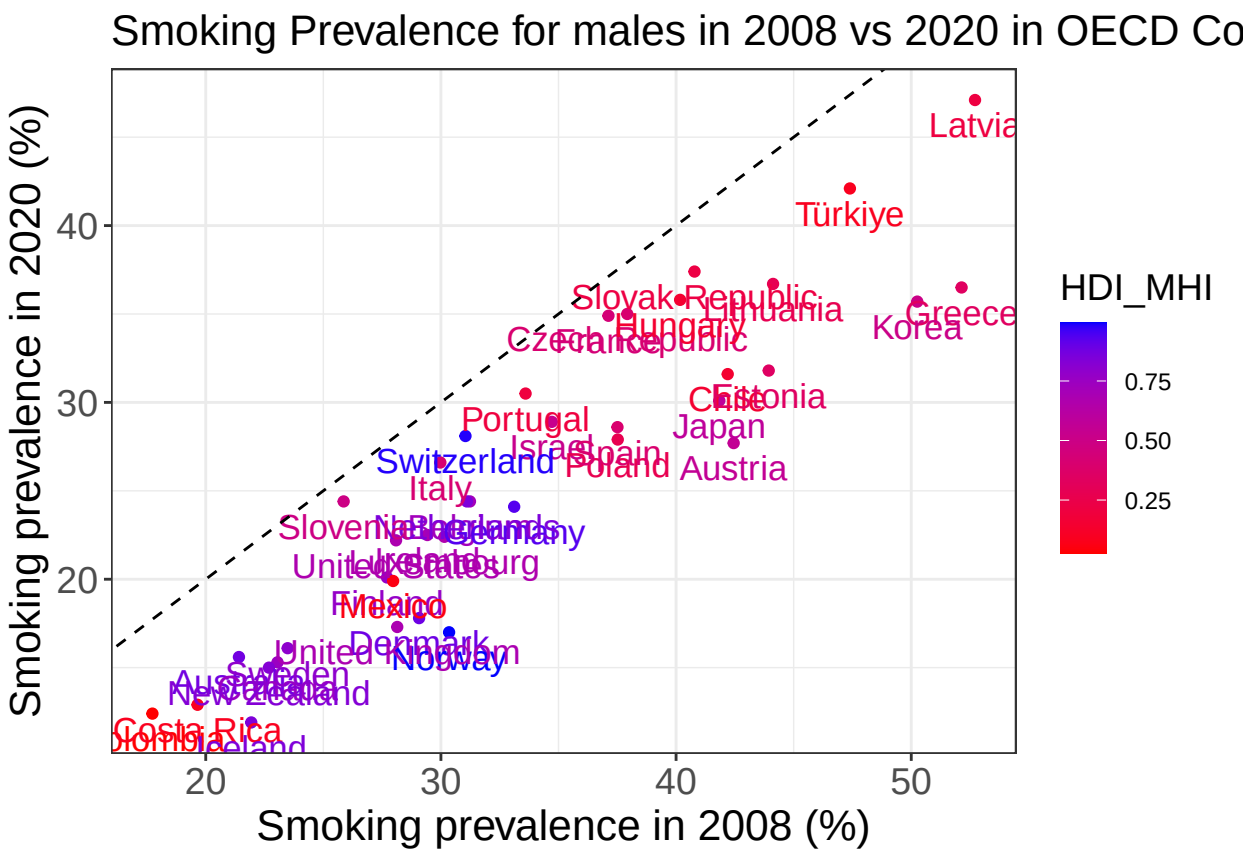
```
ggplot(data=databub,aes(x = data_2008_females,
                        y = data_2020_females,
                        color = HDI_MHI)) +
  geom_point() +
  geom_text(aes(label = countries), vjust = 1.5,size=4.5) +  # Add labels with the 'Country' variable
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "black") +  # Add the y = x line
  scale_color_gradient(low = "red", high = "blue") +
  labs(x = "Smoking prevalence in 2008 (%)",
       y = "Smoking prevalence in 2020 (%)",
       size = "Population",
       color = "HDI_MHI") +
  theme_bw() +
  theme(
    axis.title = element_text(size = 12),
    axis.text = element_text(size = 10),  # Increase the size of the axis labels
    plot.title = element_text(size = 12),
    legend.title = element_text(size = 10),
    panel.background = element_rect(fill = "white")  # Set a white color for the panel background
  ) +
```

```
  # Increase the size of the color label
  ggtitle("Smoking Prevalence for females in 2008 vs 2020 in OECD Countries")
```




We now move to males

```
ggplot(data=databub,aes(x = data_2008_males,
                        y = data_2020_males,
                        color = HDI_MHI)) +
  geom_point() +
  geom_text(aes(label = countries), vjust = 1.5,size=4.5) +  # Add labels with the 'Country' variable
  geom_abline(intercept = 0, slope = 1, linetype = "dashed", color = "black") +  # Add the y = x line
  scale_color_gradient(low = "red", high = "blue") +
  labs(x = "Smoking prevalence in 2008 (%)",
       y = "Smoking prevalence in 2020 (%)",
       size = "Population",
       color = "HDI_MHI") +
  theme_bw() +
  theme(
    axis.title = element_text(size = 16),
        axis.text = element_text(size = 14),  # Increase the size of the axis labels
        plot.title = element_text(size = 16),
        legend.title = element_text(size = 14),
    panel.background = element_rect(fill = "white")  # Set a white color for the panel background
  ) +
# Increase the size of the color label
  ggtitle("Smoking Prevalence for males in 2008 vs 2020 in OECD Countries")
```


Smoking Prevalence for males in 2008 vs 2020 in OECD Co
Smoking prevalence in 2020 (%)
Smoking prevalence in 2008 (%)
20
30
40
50
Latvia
Türkiye
Slovak Republic
Lithuania
Hungary
Greece
Korea
Czech Republic
France
Chile
Estonia
Portugal
Japan
Israel
Spain
Switzerland
Poland
Austria
Italy
Slovenia
Netherlands
Germany
United States
Luxembourg
Ireland
Finland
Mexico
Denmark
United Kingdom
Norway
Sweden
Australia
New Zealand
Costa Rica
Iceland
HDI_MHI
0.75
0.50
0.25