

## **Session3.2: Visualisation avec ggplot: Extension**

Visseho Adjiwanou, PhD.

SICSS - Montréal

09 June 2021

**Extension**

# Introduction

- Nous allons utiliser les données:
- dhs\_ipv
- crsc\_small

```
library(tidyverse)
library(ggpubr)

dhs_ipv <- read_csv("../données/dhs_ipv.csv")

dhs_ipv <-
  dhs_ipv %>%
  mutate(beat_burnfood_cat = factor(ntile(beat_burnfood, 4)),
         beat_goesout_cat = factor(ntile(beat_goesout, 4)),
         sec_school_cat = factor(ntile(sec_school, 3), labels = "1", "2", "3"),
         no_media_cat = factor(ntile(no_media, 3), labels = "1", "2", "3"))
```

# Introduction

- Plus dans aes : **mappage** : c'est une mise en relation entre un **attribut graphique du geom** et une variable du tableau de données.

# Introduction

- Plus dans aes : **mappage** : c'est une mise en relation entre un **attribut graphique du geom** et une variable du tableau de données.
- Changer les couleurs (*color*), la taille (*size*), la position (*position*), la transparence (*alpha*), le remplissage (*fill*)

# Introduction

- Plus dans aes : **mappage** : c'est une mise en relation entre un **attribut graphique du geom** et une variable du tableau de données.
- Changer les couleurs (*color*), la taille (*size*), la position (*position*), la transparence (*alpha*), le remplissage (*fill*)
- **Facets** : le **faceting** permet d'effectuer plusieurs fois le même graphique selon les valeurs d'une ou plusieurs variables qualitatives (notre *group\_by*): `facet_wrap`, `facet_grid`

# Introduction

- Plus dans aes : **mappage** : c'est une mise en relation entre un **attribut graphique du geom** et une variable du tableau de données.
- Changer les couleurs (*color*), la taille (*size*), la position (*position*), la transparence (*alpha*), le remplissage (*fill*)
- **Facets** : le **faceting** permet d'effectuer plusieurs fois le même graphique selon les valeurs d'une ou plusieurs variables qualitatives (notre *group\_by*): `facet_wrap`, `facet_grid`
- Les **scales** : ils permettent de modifier la manière dont un attribut graphique va être relié aux valeurs d'une variable, et dont la légende correspondante va être affichée.

# Introduction

- Plus dans aes : **mappage** : c'est une mise en relation entre un **attribut graphique du geom** et une variable du tableau de données.
- Changer les couleurs (*color*), la taille (*size*), la position (*position*), la transparence (*alpha*), le remplissage (*fill*)
- **Facets** : le **faceting** permet d'effectuer plusieurs fois le même graphique selon les valeurs d'une ou plusieurs variables qualitatives (notre *group\_by*): `facet_wrap`, `facet_grid`
- Les **scales** : ils permettent de modifier la manière dont un attribut graphique va être relié aux valeurs d'une variable, et dont la légende correspondante va être affichée.
- Les **thèmes** : ils permettent de contrôler l'affichage de tous les éléments du graphique qui ne sont pas reliés aux données :



## **Exemples: Visualiser la distribution univariée: extension**

## Exemples: Visualiser la distribution univariée: extension

- Plus dans aes : mappage
- Facets
- Représentation de plusieurs graphiques sur une même figure
- Les scales
- Les thèmes

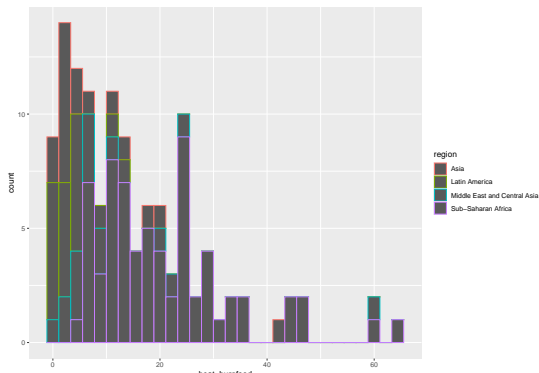
# 1. Exemple: Plus dans aes

- A la différence de boxplot qui nous permet de comparer les distributions de manière compacte entre différents groupes, les autres graphiques portaient sur l'ensemble de l'échantillon.
- Que faire pour comparer par exemple des histogrammes ou des diagrammes de barres de différents sous-échantillons?
- Vous devez penser immédiatement à :
- **by\_group**
- Malheureusement, il ne marche pas avec ggplot.
- Il faut utiliser le **mappage**
- Un mappage, dans ggplot2, est une mise en relation entre un **attribut graphique du geom** (position (*position*), couleur (*color*), taille (*size*)...) et une variable du tableau de données.

## Exemple: Plus dans aes

```
ggplot(dhs_ipv) +  
  geom_histogram(aes(x = beat_burnfood, color = region))
```

## Warning: Removed 31 rows containing non-finite values (s

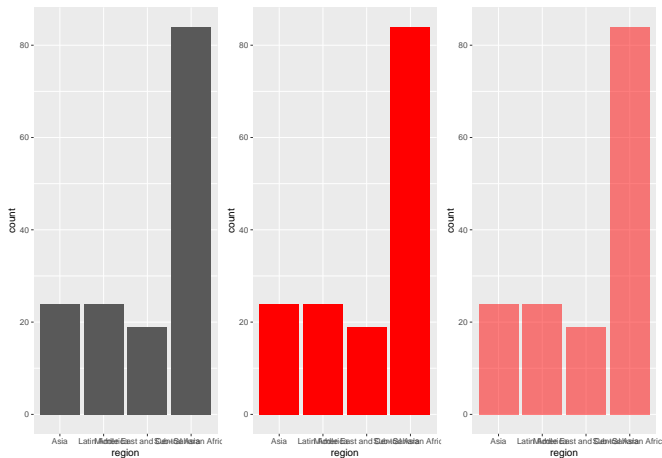


# 1. Exemple: Plus dans aes

```
c <- ggplot(dhs_ipv) +  
  geom_bar(aes(x = region))  
  
d <- ggplot(dhs_ipv) +  
  geom_bar(aes(x = region), fill = "red")  
  
e <- ggplot(dhs_ipv) +  
  geom_bar(aes(x = region), fill = "red", alpha = 0.5)
```

# 1. Exemple: Plus dans aes

```
ggarrange(c, d, e, ncol = 3)
```



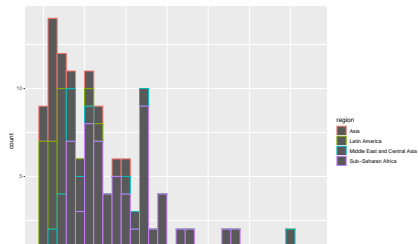
## Exemple: Facets

## 2. Exemple: Facets

- Le **faceting** permet d'effectuer plusieurs fois le même graphique selon les valeurs d'une ou plusieurs variables qualitatives (notre *group\_by*).

```
ggplot(dhs_ipv) +  
  geom_histogram(aes(x = beat_burnfood, color = region))
```

## Warning: Removed 31 rows containing non-finite values (s

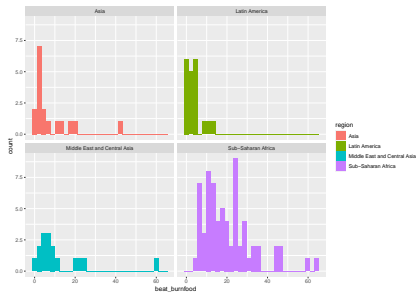




## 2. Exemple: Facets

```
ggplot(dhs_ipv) +  
  geom_histogram(aes(x = beat_burnfood, fill = region)) +  
  facet_wrap(~region)
```

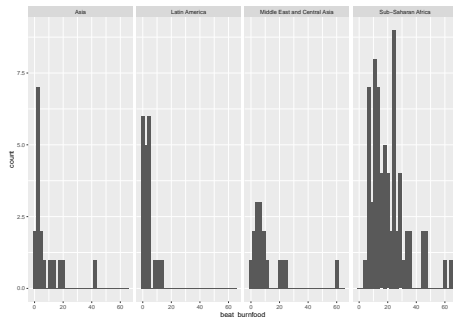
## Warning: Removed 31 rows containing non-finite values (s



## 2. Exemple: Facets

```
ggplot(dhs_ipv) +  
  geom_histogram(aes(x = beat_burnfood)) +  
  facet_grid(~region)
```

## Warning: Removed 31 rows containing non-finite values (s



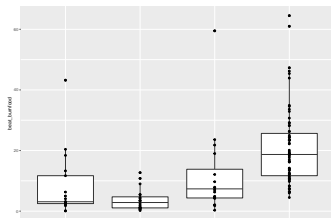
## Représentation de plusieurs graphiques sur une même figure

## Représentation de plusieurs graphiques sur une même figure

```
ggplot(dhs_ipv) +  
  geom_boxplot(aes(x = region, y = beat_burnfood)) +  
  geom_point(aes(x = region, y = beat_burnfood))
```

```
## Warning: Removed 31 rows containing non-finite values (stats::boxplot.outliers)
```

```
## Warning: Removed 31 rows containing missing values (geom_point)
```



## Représentation de plusieurs graphiques sur une même figure

- Vous pouvez observer qu'ici, on a répété l'information dans le aes dans les deux geom.
- On peut faire plus simple

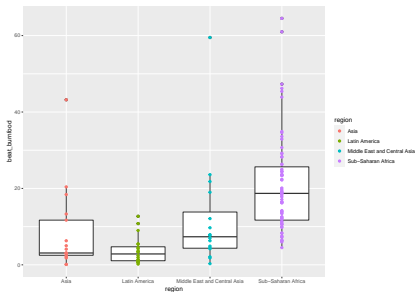
```
a <- ggplot(dhs_ipv) +  
  geom_boxplot(aes(x = region, y = beat_burnfood)) +  
  geom_point(aes(x = region, y = beat_burnfood))  
  
b <- ggplot(dhs_ipv, aes(x = region, y = beat_burnfood)) +  
  geom_boxplot() +  
  geom_point(aes(color = region))
```

# Représentation de plusieurs graphiques sur une même figure

b

```
## Warning: Removed 31 rows containing non-finite values (s
```

```
## Warning: Removed 31 rows containing missing values (geon
```



## Les scales

# Les scales

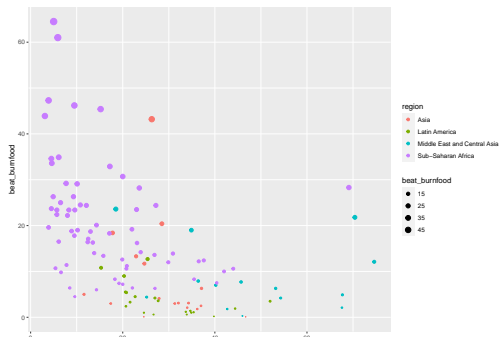
- Les scales dans ggplot2 permettent de modifier la manière dont un attribut graphique va être relié aux valeurs d'une variable, et dont la légende correspondante va être affichée.
- Par exemple, pour l'attribut `color`, on pourra définir la palette de couleur utilisée. Pour `size`, les tailles minimales et maximales, etc.



# Les scales

```
ggplot(dhs_ipv) +
  geom_point(aes(x = sec_school, y = beat_burnfood, color =
    scale_size(range = c(0,4), breaks = c(15, 25, 35, 45))
```

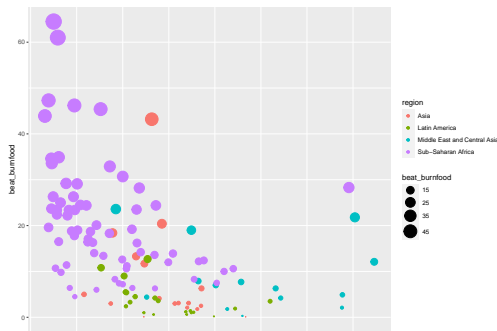
## Warning: Removed 33 rows containing missing values (geom



# Les scales

```
ggplot(dhs_ipv) +
  geom_point(aes(x = sec_school, y = beat_burnfood, color =
    scale_size(range = c(0,10), breaks = c(15, 25, 35, 45))
```

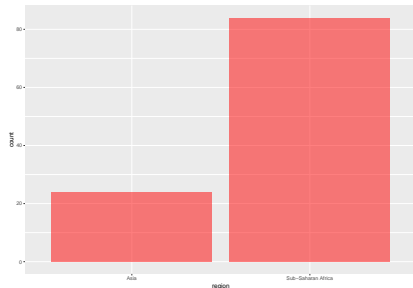
## Warning: Removed 33 rows containing missing values (geom



## Les scales

```
ggplot(dhs_ipv) +  
  geom_bar(aes(x = region), fill = "red", alpha = 0.5) +  
  scale_x_discrete("region", limits = c("Asia", "Sub-Saharan Africa"))
```

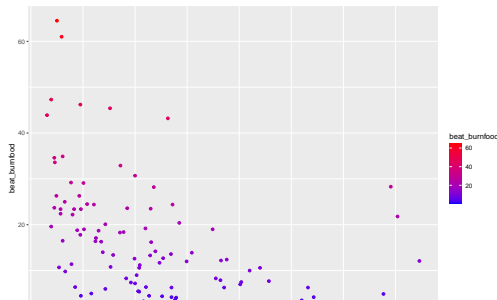
## Warning: Removed 43 rows containing non-finite values (s



# Les scales

```
ggplot(dhs_ipv) +  
  geom_point(aes(x = sec_school, y = beat_burnfood, color =  
    #geom_text(aes(x = sec_school, y = beat_burnfood, label =  
    scale_color_gradient("beat_burnfood", low = "blue", high =
```

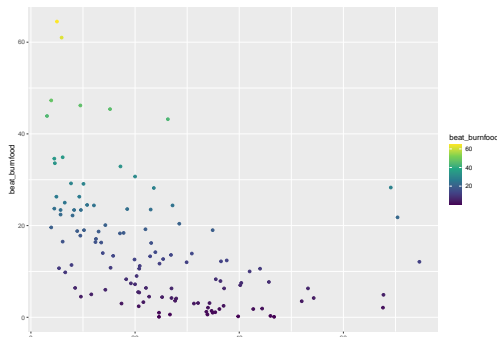
## Warning: Removed 33 rows containing missing values (geom



# Les scales

```
ggplot(dhs_ipv) +  
  geom_point(aes(x = sec_school, y = beat_burnfood, color =  
    scale_color_viridis_c("beat_burnfood") # option = "inferno"
```

## Warning: Removed 33 rows containing missing values (geom



## Les thèmes

# Les thèmes

<https://ggplot2.tidyverse.org/reference/theme.html>

- Les thèmes permettent de contrôler l'affichage de tous les éléments du graphique qui ne sont pas reliés aux données : titres, grilles, fonds, etc.

```
a <- ggplot(dhs_ipv) +  
  geom_histogram(aes(x = beat_burnfood)) +  
  ggtitle("Histogramme") +  
  xlab("Proportion des femmes qui estiment qu'une femme doit  
  ylab("Effectif") +  
  theme_bw()
```

# Les thèmes

a

```
## Warning: Removed 31 rows containing non-finite values (s
```

