

# Burger

Sefkan Lamia Chakir

27/10/2021

## Présentation du jeu de données

Pour commencer, notre jeu de données est constitué de 80 items pour 14 variables.

```
dim(burger)
```

```
## [1] 80 14
```

Nous avons des variables sous différentes unités de mesures : le cholestérol et le sodium sont en milligramme alors que le reste des macronutriments sont en gramme, nous avons donc besoin de nettoyer les données afin de travailler sur le même ordre de grandeur pour ces variables.

Nous avons trois restaurants qui proposent différents type d'item (i.e "articles") ainsi que des formules pour le petit-déjeuner, voici le nombre de formules petit-déjeuner proposer par chaque restaurants :

```
##
##      BurgerKing McDonalds Wendys
## No           24           25      28
## Yes           0            3       0
```

Cependant McDonalds est le seul des 3 restaurants à proposer des formules petit-déjeuner, c'est pourquoi nous décidons de les retirer car les données ne seront pas suffisante pour en tirer des informations utiles. Nous avons donc maintenant 77 items.

Voici la répartition des items pour chaque restaurant :

```
##
##      BurgerKing McDonalds Wendys
## Dessert         2          3      0
## Other           1          3      2
## Salad           0          3      6
## Sandwich        15         10     13
## Side            6          1      4
## Wrap            0          5      3
```

Nous pouvons nous rendre compte que la majorité des items sont des Sandwich.

## Quel restaurant est le plus healthy ?

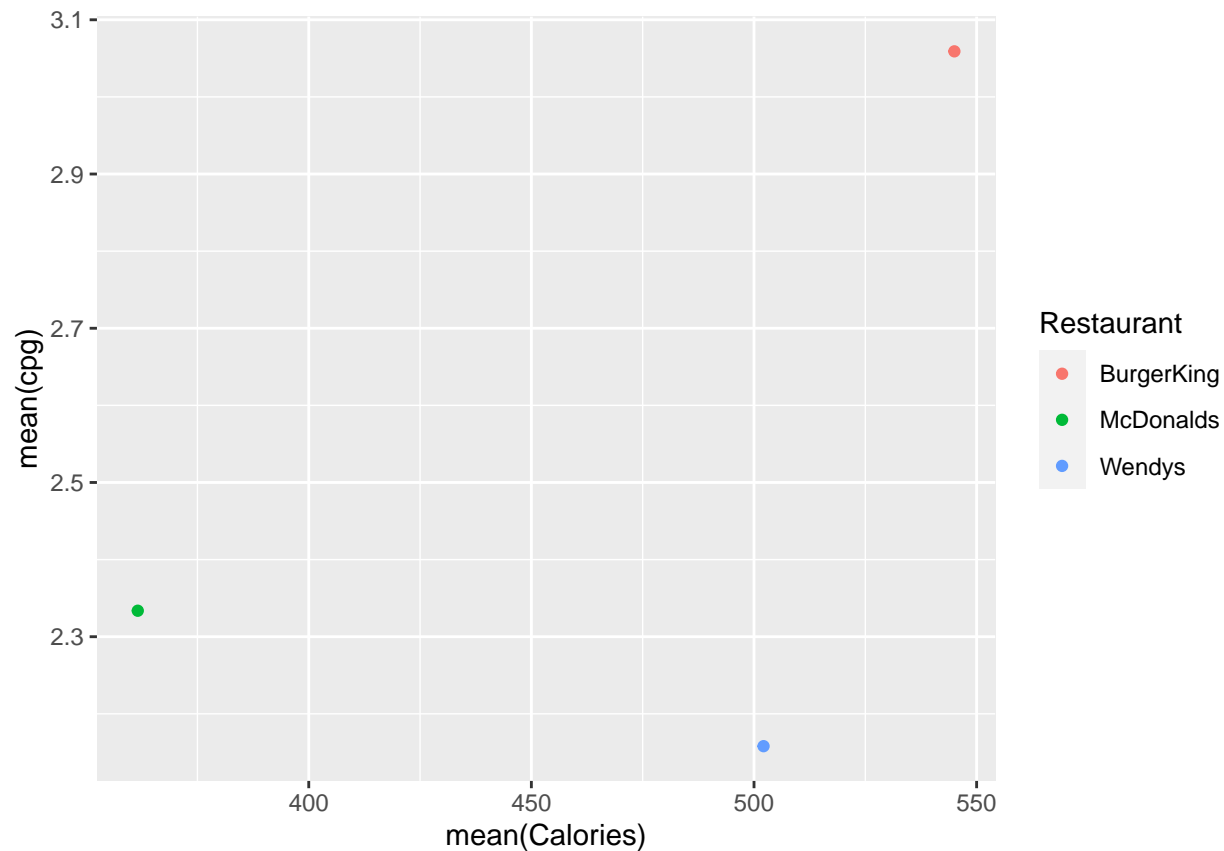
- Comparer la moyenne calorique de chaque restaurant

```
## # A tibble: 3 x 2
##   Restaurant 'mean(Calories)'
##   <fct>          <dbl>
## 1 BurgerKing      545
## 2 McDonalds      362.
## 3 Wendys         502.
```

Ici nous observons que McDonalds est en moyenne plus calorique que les deux autres restaurants, nous pouvons nous demander si la taille de la portion peut en être la cause. Pour vérifier cela nous allons calculer la moyenne de calorique de chaque gramme pour chaque item d'un restaurant, pour faire simple, nous répondons à la question : Combien de calorie il y a en moyenne dans chaque gramme des items d'un restaurant. Nous trouvons le résultat suivant :

```
## # A tibble: 3 x 2
##   Restaurant 'mean(cpg)'
##   <fct>          <dbl>
## 1 BurgerKing      3.06
## 2 McDonalds      2.33
## 3 Wendys         2.16
```

```
##   Restaurant mean(Calories) mean(cpg)
## 1 BurgerKing      545.0000  3.059078
## 2 McDonalds      361.6000  2.333700
## 3 Wendys         502.1429  2.158076
```

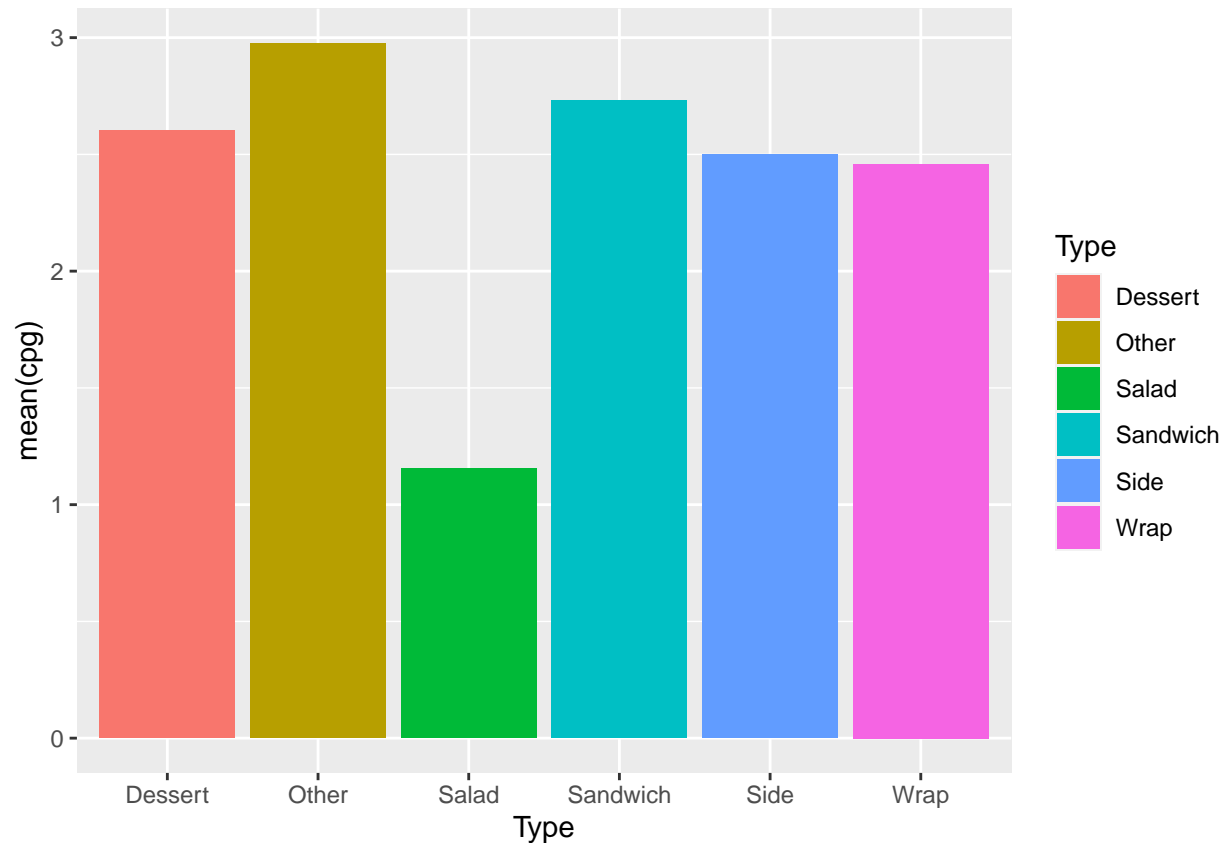


Ici nous voyons donc que la moyenne des calories par gramme d'un restaurant n'a pas de rapport avec la moyenne des calories d'un restaurant. A moins que cette différence vienne d'autre part, notamment du type des items.

Pour cela nous pouvons commencer par regarder quelles types d'items sont les plus caloriques, et lesquels sont les moins caloriques :

```
## # A tibble: 6 x 2
##   Type      'mean(cpg)'
##   <fct>      <dbl>
## 1 Dessert      2.60
## 2 Other        2.98
## 3 Salad        1.16
## 4 Sandwich     2.73
## 5 Side         2.50
## 6 Wrap         2.46

## Warning: Ignoring unknown parameters: binwidth, bins, pad
```



Nous observons que les salades sont de loin le type le moins calorique, ils possèdent moins de la moitié des grammes par rapport aux autres types.

Sur la table des type/restaurant nous voyons que BurgerKing, le restaurant le plus calorique, ne propose aucune salade dans son menu. Nous allons donc regarder ce que donne le resultat des calories par gramme en retirant les salades de la liste.

```
## # A tibble: 3 x 2
##   Restaurant 'mean(cpg)'
##   <fct>         <dbl>
## 1 BurgerKing     3.06
## 2 McDonalds     2.58
## 3 Wendys        2.34
```

Nous voyons que l'écart se réduit mais BurgerKing reste toujours le restaurant le plus calorique. Afin de comprendre pourquoi nous allons comparer chaque type de chaque restaurant pour comprendre d'où cette différence peut venir.

```
mcdo = burger %>% filter(Restaurant=="McDonalds")
wendys = burger %>% filter(Restaurant=="Wendys")
bk = burger %>% filter(Restaurant=="BurgerKing")

mcdoCpgType = mcdo %>% group_by(Restaurant) %>% group_by(Type) %>% summarise(mean(cpg))
wendysCpgType = wendys %>% group_by(Restaurant) %>% group_by(Type) %>% summarise(mean(cpg))
bkCpgType = bk %>% group_by(Restaurant) %>% group_by(Type) %>% summarise(mean(cpg))

mcdoCpgType$Restaurant = as.factor("McDonalds")
```

```
wendysCpgType$Restaurant = as.factor("Wendys")
bkCpgType$Restaurant = as.factor("BurgerKing")
mcdoCpgType
```

```
## # A tibble: 6 x 3
##   Type      'mean(cpg)' Restaurant
##   <fct>      <dbl> <fct>
## 1 Dessert      2.03 McDonalds
## 2 Other        3.06 McDonalds
## 3 Salad        0.500 McDonalds
## 4 Sandwich     2.60 McDonalds
## 5 Side         3.25 McDonalds
## 6 Wrap         2.47 McDonalds
```

```
#cpgTypeRests$BurgerKin =
```

## Ce qu'il y a en dessous c'est du brouillon

Effectif croisés

```
aa = table(burger$Type, burger$Restaurant)
table(burger$Breakfast, burger$Restaurant)
```

```
##
##      BurgerKing McDonalds Wendys
## No           24          25      28
## Yes           0           0       0
```

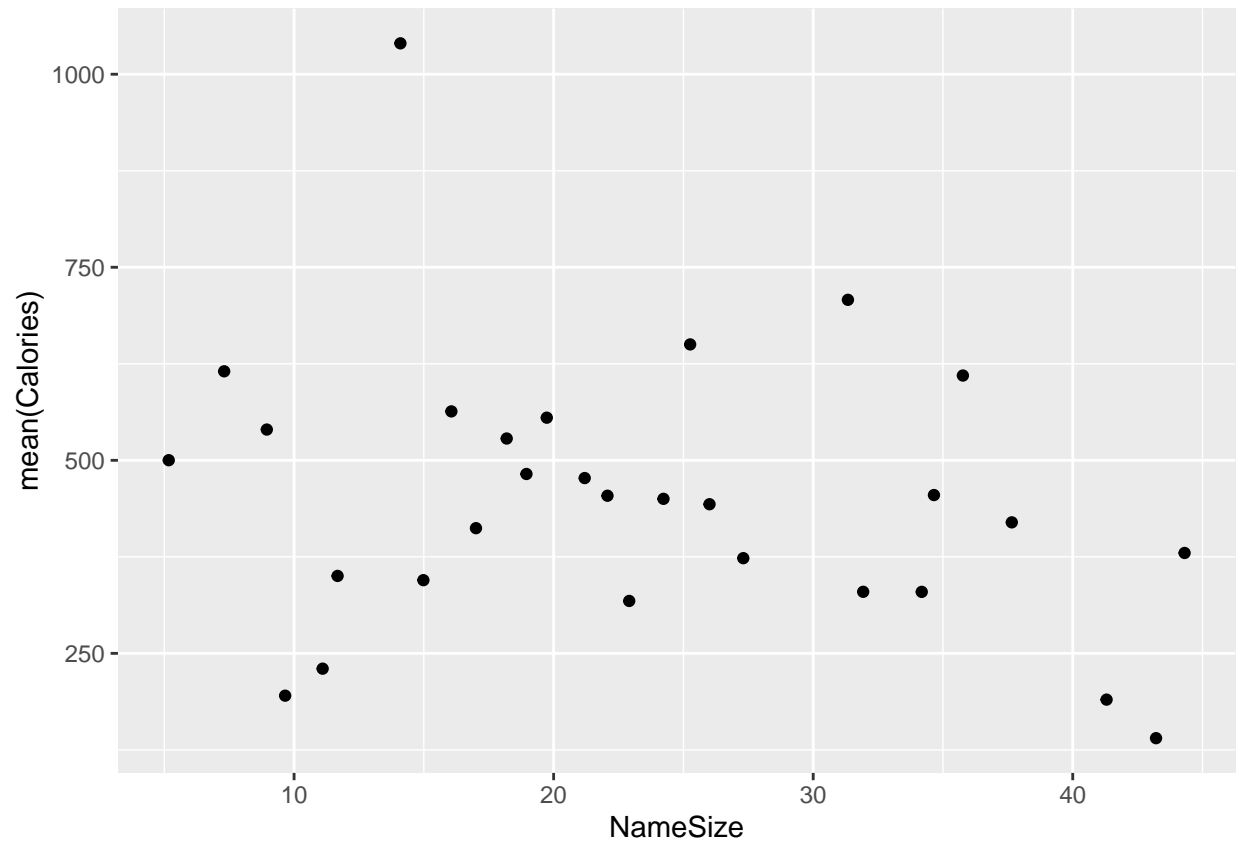
```
aa
```

```
##
##      BurgerKing McDonalds Wendys
## Dessert         2          3       0
## Other           1          3       2
## Salad           0          3       6
## Sandwich        15         10      13
## Side            6          1       4
## Wrap            0          5       3
```

```
#aa %>% corrplot()
```

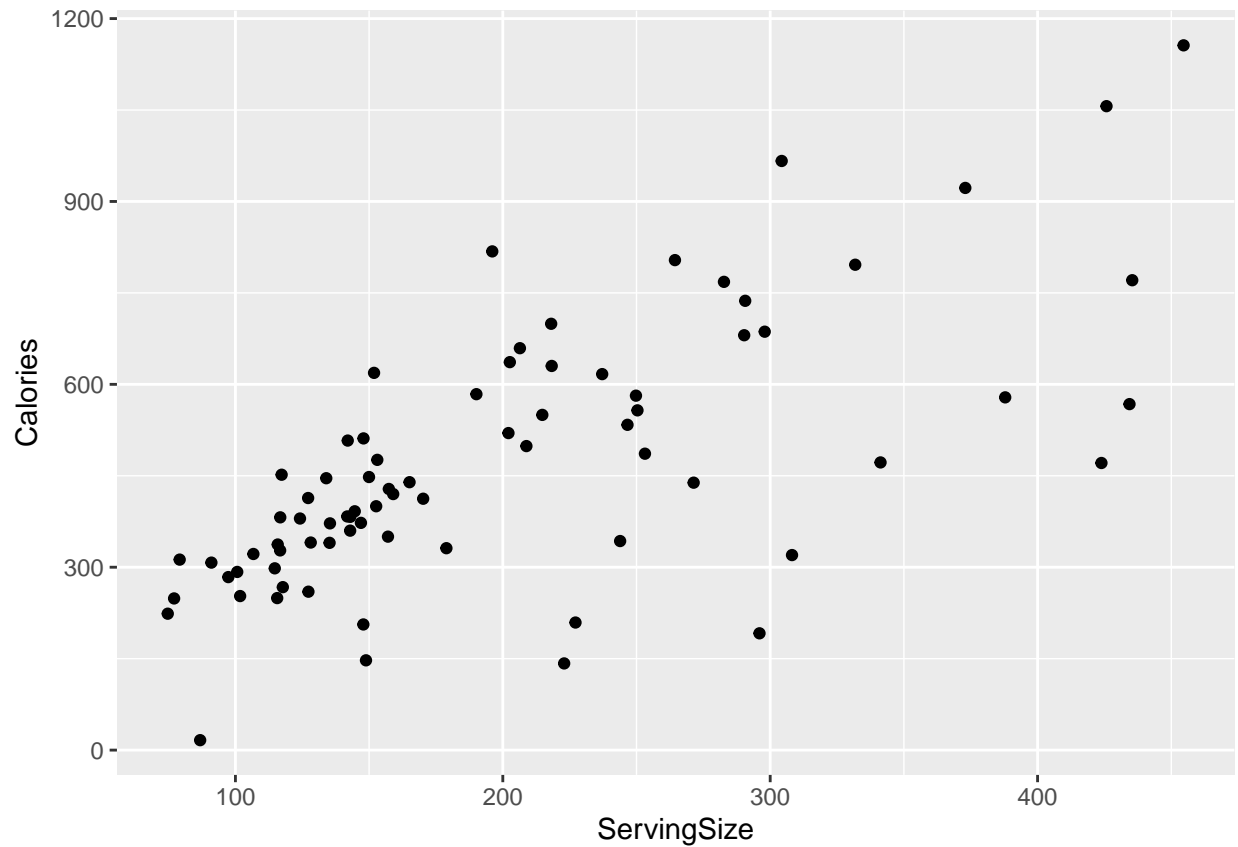
Analyse des calories en fonction de la taille du nom de l'item

```
###A revoir plus tard
bgNameSize = burger %>% mutate(NameSize=nchar(Item))
bgNameSizeAndCal = bgNameSize %>% group_by(NameSize) %>% summarise(mean(Calories)) %>% arrange(`mean(Calories)`)
bgNameSizeAndCal %>% ggplot(aes(x=NameSize, y=`mean(Calories)`) + geom_jitter())
```



Lien entre Serving Size et calories

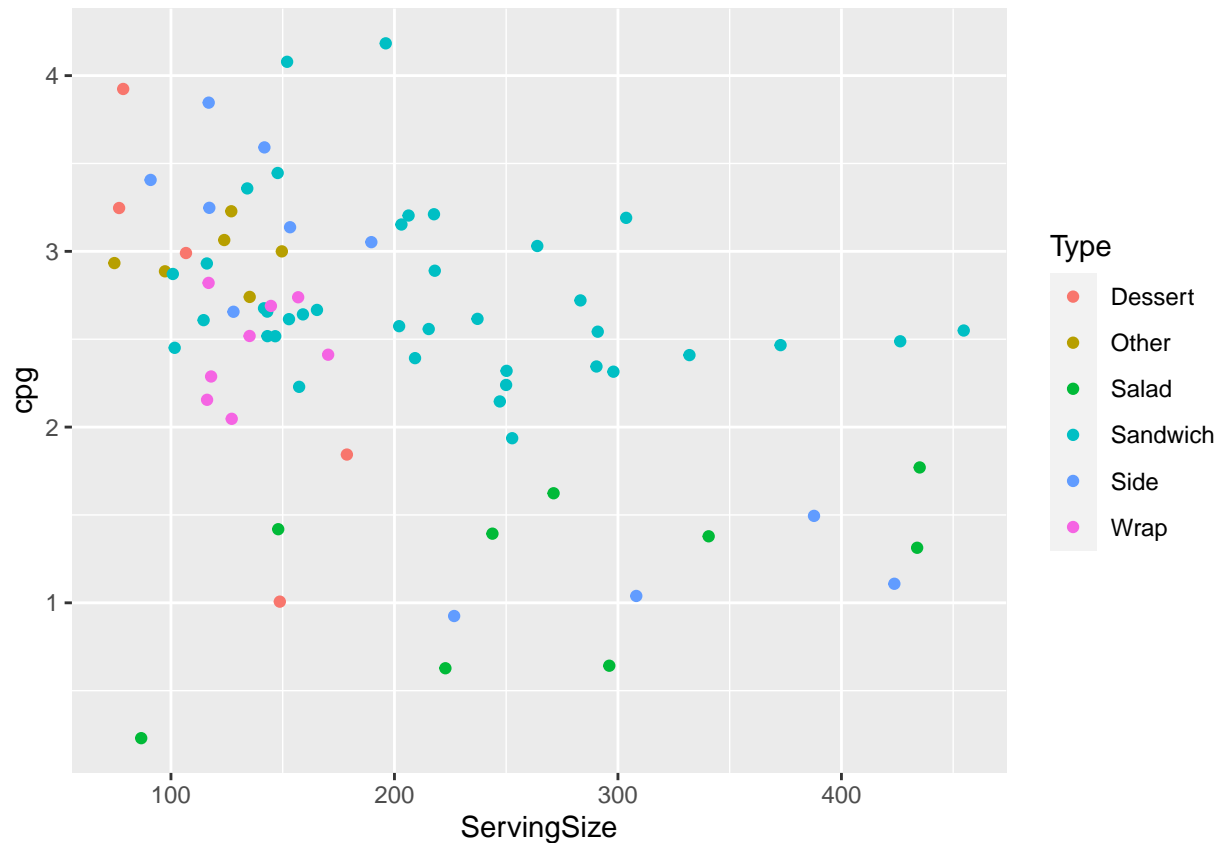
```
burger %>% ggplot(aes(x=ServingSize, y=Calories)) + geom_jitter()
```



On remarque que plus la quantité est grande plus le nombre de calories est grand. On se demande donc quel peut être le résultat en faisant ça avec le nombre de calorie par gramme.

serving size divisé par calorie cpg = calorie per gram

```
burger$cpg = burger$Calories / burger$ServingSize  
burger %>% ggplot(aes(x=ServingSize, y=cpg, color=Type)) + geom_jitter()
```



On remarque donc qu'il n'y a pas de rapport entre la quantité par item et son nombre de calories.

On peut toutefois remarquer que certain type d'item ont des quantités calorique similaire.

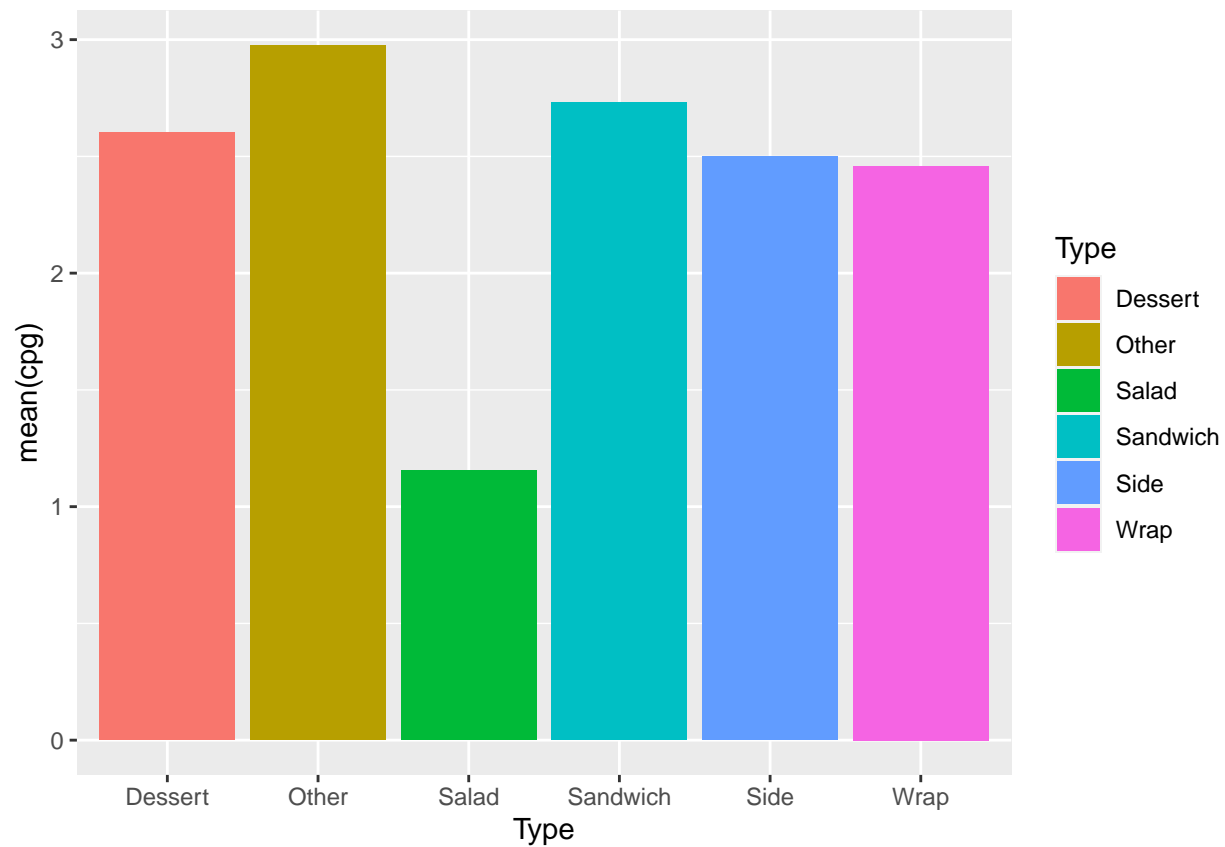
```
caloriesPerType = burger %>% group_by(Type) %>% summarise(mean(cpg))
caloriesPerType
```

```
## # A tibble: 6 x 2
##   Type      'mean(cpg)'
##   <fct>      <dbl>
## 1 Dessert      2.60
## 2 Other        2.98
## 3 Salad        1.16
## 4 Sandwich     2.73
## 5 Side         2.50
## 6 Wrap         2.46
```

```
caloriesPerType %>% ggplot(aes(x=Type, y=`mean(cpg)`, fill=Type)) + geom_histogram(stat="identity")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```





Moyenne de la taille des plats servis par restaurant Et moyenne des calories de chaque plat

```
#burger %>% group_by(Restaurant) %>% summarise(mean = mean(ServingSize), n = n())
#burger %>% group_by(Restaurant) %>% summarise("moyenne (kcal)" = mean(Calories), n = n())
```