

Le pouvoir d’achat des ouvriers anglais du XVIe au XIXe siècle - Exercice évalué par les pairs

2023-05-09

Contexte

William Playfair était un des pionniers de la présentation graphique des données. Il est notamment considéré comme l'inventeur de l'histogramme. Un de ses graphes célèbres, tiré de son livre "A Letter on Our Agricultural Distresses, Their Causes and Remedies", montre l'évolution du prix du blé et du salaire moyen entre 1565 et 1821. Playfair n'a pas publié les données numériques brutes qu'il a utilisées, car à son époque la répliquabilité n'était pas encore considérée comme essentielle.

Importation des données

Les données brutes ont été obtenues après numérisation du graphe, et sont aujourd'hui téléchargeables [ici](#)

Ici, on a téléchargé le fichier de données en local, on peut donc le lire, et regarder sa structure avec la fonction `str()`

```
data <- read.csv("/Volumes/SD Laurie/Documents/THÈSE/Formations/MOOC Recherche Reproducible/Wheat.csv")
str(data)

## 'data.frame':    53 obs. of  4 variables:
## $ X      : int   1  2  3  4  5  6  7  8  9 10 ...
## $ Year    : int 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 ...
## $ Wheat   : num  41 45 42 49 41.5 47 64 27 33 32 ...
## $ Wages   : num  5 5.05 5.08 5.12 5.15 5.25 5.54 5.61 5.69 5.78 ...
```

On constate que les années sont des entiers, tandis que le prix du blé et des salaires sont en format numérique.

On peut également vérifier si le jeu de données contient des données manquantes :

```
which(is.na(data), arr.ind=TRUE)

##      row col
## [1,]   51   4
## [2,]   52   4
## [3,]   53   4
```

Il y a 3 données manquantes, qui correspondent au salaire hebdomadaire (colonne n°4) des trois dernières lignes, c'est-à-dire des années 1815, 1820, 1821.

Objectif n°1 : reproduire le graphe de Playfair

La consigne est la suivante : "Votre première tâche est de reproduire le graphe de Playfair à partir des données numériques. Représentez, comme Playfair, le prix du blé par des barres et les salaires par une surface bleue délimitée par une courbe rouge. Superposez les deux de la même façon dans un seul graphique. Le style de votre graphique pourra rester différent par rapport à l'original, mais l'impression globale devrait être la même."

Téléchargement des packages nécessaires: ggplot2 pour les graphiques

```
library(ggplot2)

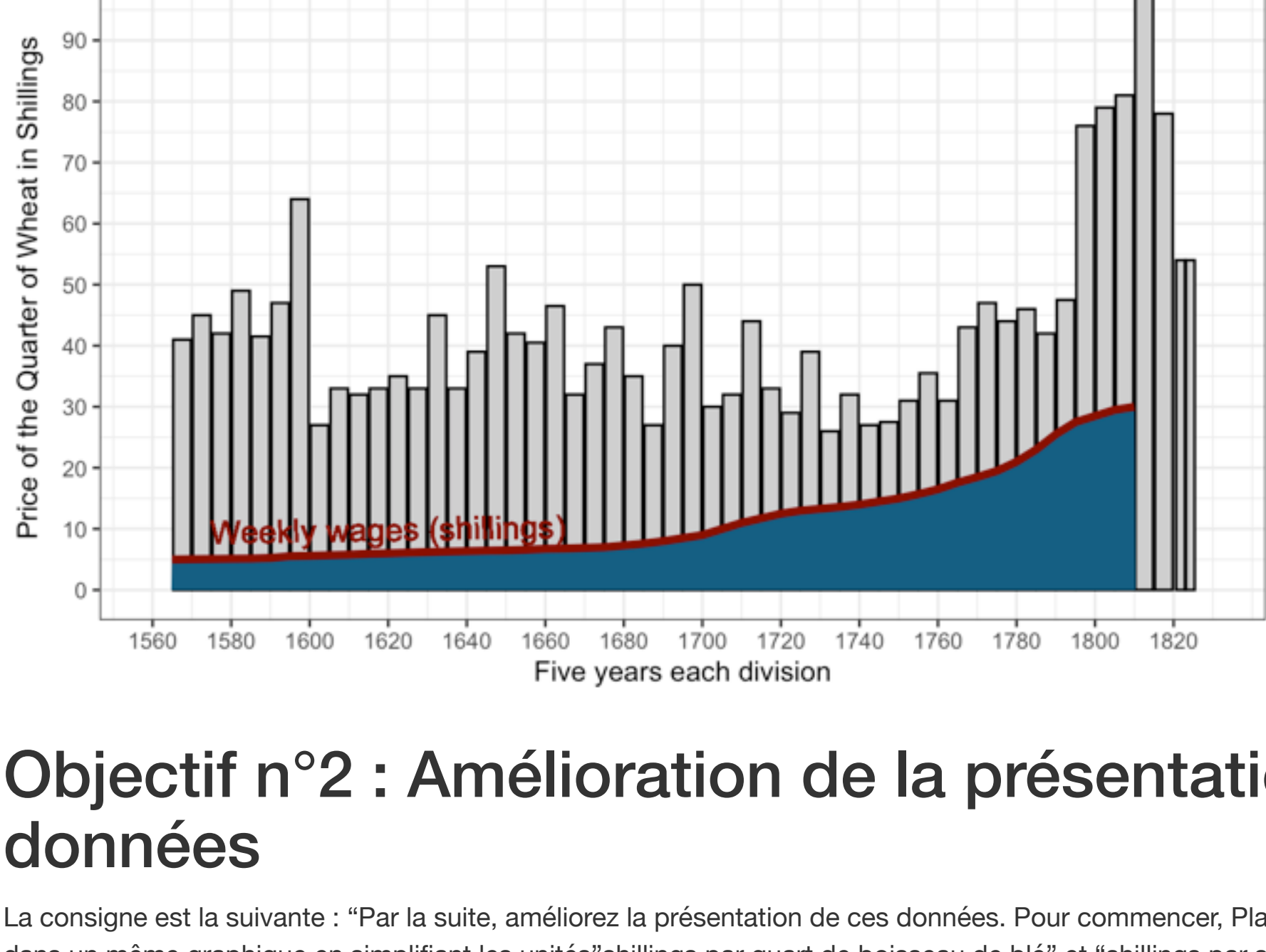
graph_playfair = ggplot(data)+
  geom_col(aes(x=Year, y=Wheat), width=5, fill="gray81", colour="black", just=0, position= position_dodge2(preserve="total"))+
  geom_area(aes(x=Year, y=Wages), color="red4", fill="#155F83FF", size=1.5)+
  scale_y_continuous(breaks=c(0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100))+
  scale_x_continuous(breaks=seq(1560, 1830, 20), labels=seq(1560, 1830, 20), limits=c(1560, 1830))
```

On ajoute le titre du graphique, ainsi que celui de l'axe des ordonnées et des abscisses, grâce à `labs()`. On ajoute également une petite annotation pour indiquer ce que représente la courbe rouge.

```
graph_playfair = graph_playfair +
  labs(title="Showing at One View \nthe Price of the Quarter of Wheat \n& Wages of Labor by the Week \nfrom the Year 1565 to 1821 by William Playfair", x="Five years each division", y="Price of the Quarter of Wheat in Shilling s") +
  geom_text(x=1620, y=10, label="Weekly wages (shillings)", color="red4", size=4.5, angle=1)
```

Enfin, quelques éléments graphiques (la couleur du fond du graphique notamment) :

```
graph_playfair = graph_playfair +
  theme_bw()+
  theme(legend.position="none")
graph_playfair
```



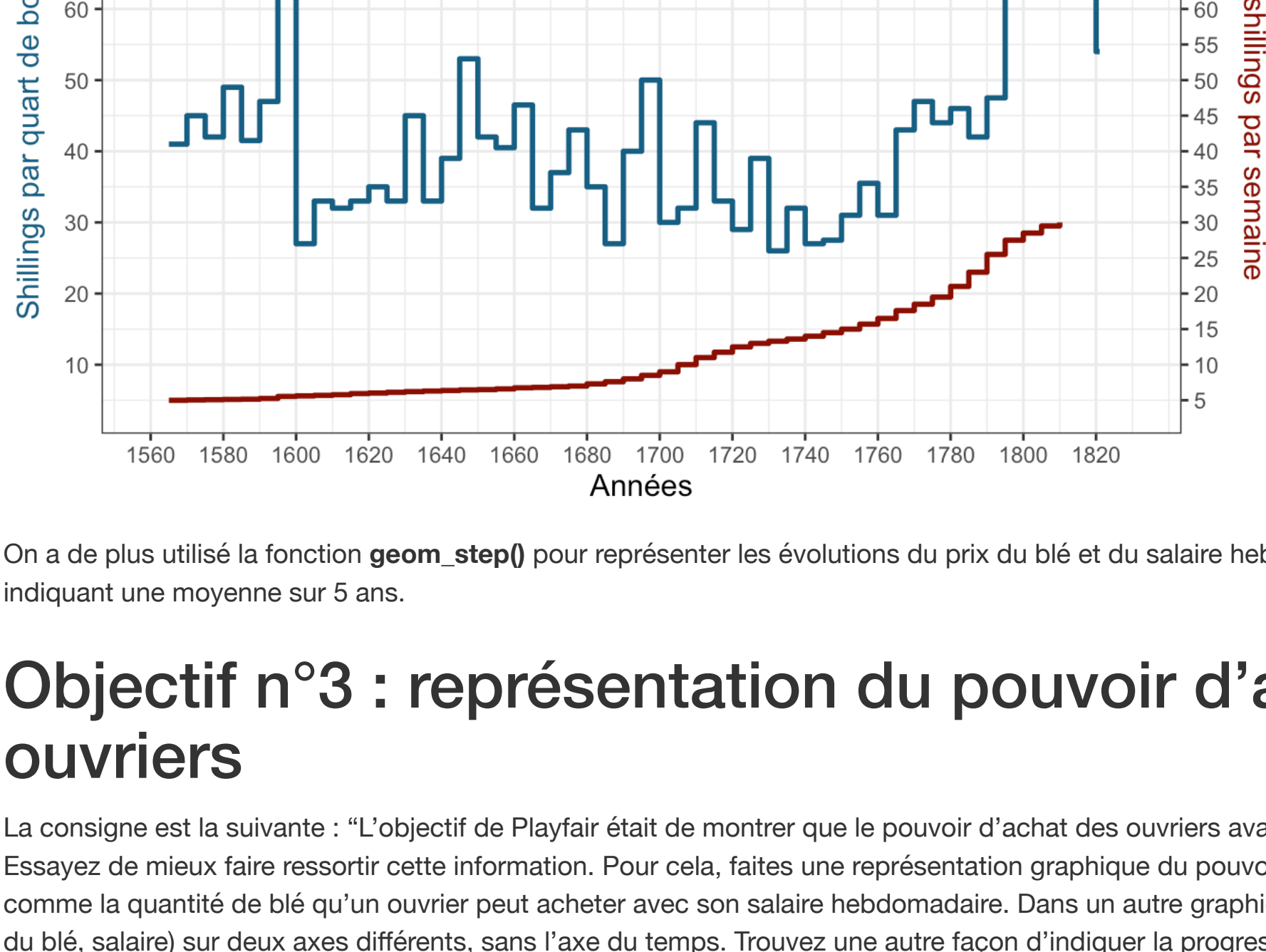
Objectif n°2 : Amélioration de la présentation des données

La consigne est la suivante : "Par la suite, améliorez la présentation de ces données. Pour commencer, Playfair a combiné les deux quantités dans un même graphique en simplifiant les unités "shillings par quart de boisseau de blé" et "shillings par semaine" à un simple "shillings", ce qui aujourd'hui n'est plus admissible. Utilisez deux ordonnées différentes, une à gauche et une à droite, et indiquez les unités correctes. À cette occasion, n'hésitez pas à proposer d'autres représentations que des barres et des surface/courbes pour les deux jeux de données si ceci vous paraît judicieux."

Changement des ordonnées

On rajoute un axe des ordonnées à droite en utilisant l'argument 'seq.axis' dans la fonction `scale_y_continuous()`. De plus, `theme()` nous permet d'indiquer des détails des 3 titres des axes, notamment la couleur, la police et la taille du texte.

```
graph_playfair2 = ggplot(data)+
  geom_step(aes(x=Year, y=Wheat), color="#155F83FF", size=1)+
  geom_step(aes(x=Year, y=Wages), color="red4", size=1)+
  scale_x_continuous(breaks=seq(1560, 1830, 20), labels=seq(1560, 1830, 20), limits=c(1560, 1830), name="Années") +
  scale_y_continuous(breaks=seq(0,100,10), name="Shillings par quart de boisseau de blé", sec.axis = sec_axis(~./1, name="Salaire en shillings par semaine", breaks=seq(0, 100, 5)))+
  theme_bw()+
  theme(
    axis.title.y = element_text(color="#155F83FF", size=12),
    axis.title.y.right = element_text(color="red4", vjust=1, size=12),
    axis.title.x = element_text(size=12)
  )+
  ggtitle("Amélioration du graphique de Playfair, représentant \n l'évolution du prix du blé et du salaire moyen entre 1565 et 1821")
graph_playfair2
```



On a de plus utilisé la fonction `geom_step()` pour représenter les évolutions du prix du blé et du salaire hebdomadaire, chaque "marche" indiquant une moyenne sur 5 ans.

Objectif n°3 : représentation du pouvoir d'achat des ouvriers

La consigne est la suivante : "L'objectif de Playfair était de montrer que le pouvoir d'achat des ouvriers avait augmenté au cours du temps. Essayez de mieux faire ressortir cette information. Pour cela, faites une représentation graphique du pouvoir d'achat au cours du temps, définie comme la quantité de blé qu'un ouvrier peut acheter avec son salaire hebdomadaire. Dans un autre graphique, montrez les deux quantités (prix du blé, salaire) sur deux axes différents, sans l'axe du temps. Trouvez une autre façon d'indiquer la progression du temps dans ce graphique. Quelle représentation des données vous paraît la plus claire ? N'hésitez pas à en proposer d'autres."

Calcul du pouvoir d'achat

Le prix du blé dans les données est en shillings par quart de boisseau de blé, c'est-à-dire en shillings pour 6,8kg de blé.

On peut donc calculer la quantité de blé que peut acheter un ouvrier par semaine avec son salaire :

```
pouvoir_achat = (data$Wages*6.8) / data$Wheat
# On peut regarder rapidement la moyenne de cette variable
mean(pouvoir_achat, na.rm=TRUE)

## [1] 1.901199
```

En moyenne entre 1565 et 1821, un ouvrier anglais pouvait acheter 1,9kg de blé avec son salaire hebdomadaire.

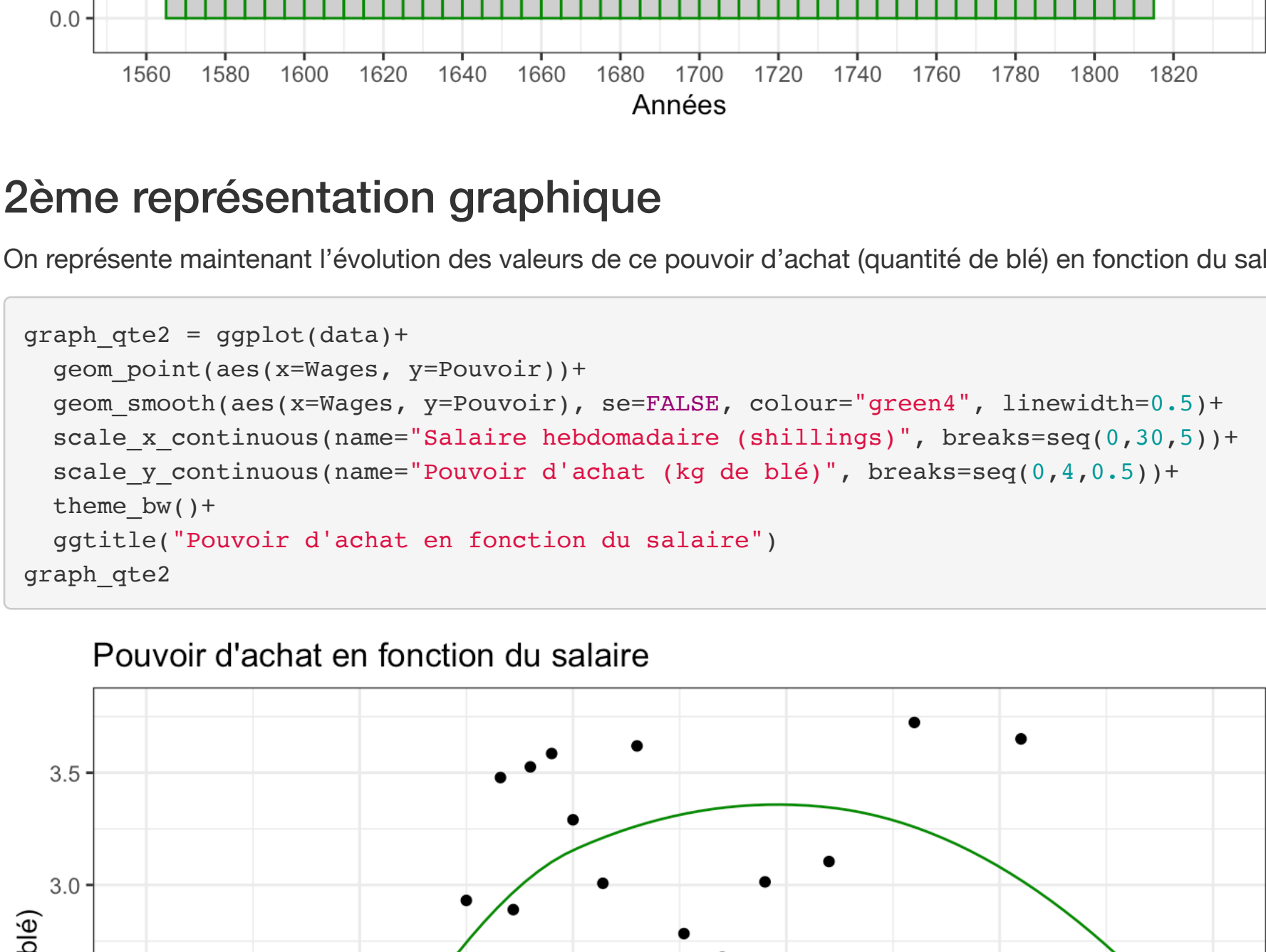
Ajoutons cette variable à notre jeu de données (pouvoir_achat est en format numérique) :

```
data$pouvoir = pouvoir_achat
```

1ère représentation graphique

On utilise la fonction `geom_col()`, pour laquelle chaque barre représente la moyenne sur 5 ans. On spécifie les limites et l'échelle de l'axe des ordonnées, ainsi que son titre. Idem pour l'axe des abscisses. On ajoute quelques éléments graphiques avec `theme_bw()`, puis le titre du graphique avec `ggtitle()`.

```
graph_gte = ggplot(data)+
  geom_col(aes(x=Year, y=Pouvoir), width=5, just=0, color="green4", fill="grey81")+
  geom_smooth(aes(x=Year, y=Pouvoir), se=FALSE, linewidth=0.5, color="green4")+
  scale_x_continuous(breaks=seq(0,4,0.5), name="Pouvoir d'achat (quantité de blé en kg)") +
  scale_y_continuous(breaks=seq(1560, 1830, 20), labels=seq(1560, 1830, 20), limits=c(1560, 1830), name="Années")
+
  theme_bw()+
  ggtitle("Pouvoir d'achat des ouvriers anglais au cours du temps")
graph_gte
```



2ème représentation graphique

On représente maintenant l'évolution des valeurs de ce pouvoir d'achat (quantité de blé) en fonction du salaire sur la période temporelle étudiée.

```
graph_gte2 = ggplot(data)+
  geom_point(aes(x=Wages, y=Pouvoir))+
  geom_smooth(aes(x=Wages, y=Pouvoir), se=FALSE, colour="green4", linewidth=0.5)+
  scale_x_continuous(name="Salaire hebdomadaire (shillings)", breaks=seq(0,30,5))+
  scale_y_continuous(name="Pouvoir d'achat (kg de blé)", breaks=seq(0,4,0.5))+
  theme_bw()+
  ggtitle("Pouvoir d'achat en fonction du salaire")
graph_gte2
```



Ajout de la progression du temps sur ce graphique

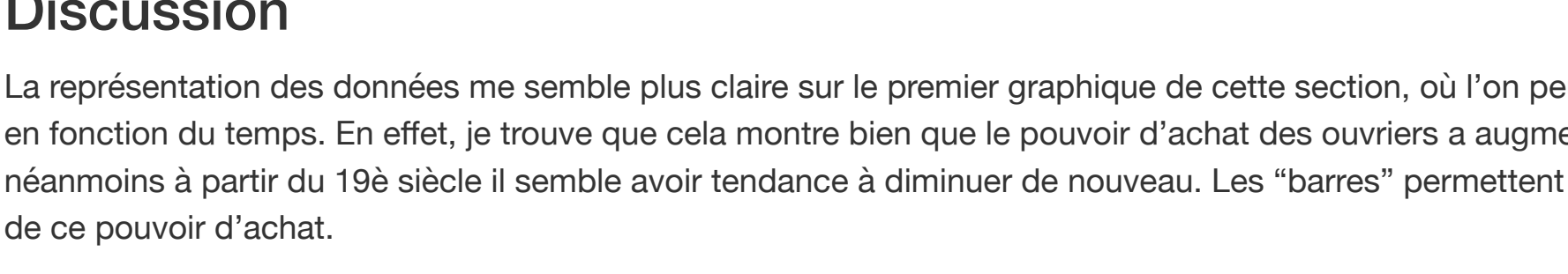
Rajoutons la progression du temps dans ce graphique : Les données s'étalant sur plusieurs siècles, il peut être pertinent de colorer les points en fonction du siècle (16^e, 17^e, 18^e et 19^e siècles).

Créons une variable "siècle" que l'on ajoute au jeu de données : (Nous avons besoin de la fonction `paste0()` qui se trouve dans la bibliothèque "string")

```
library(stringr)
data$Siècle = paste0(as.numeric(str_sub(data$Year, start=1, end=2))+1, "è")

On peut maintenant ajouter cette variable sur notre graphique (en reprenant les points du graphique précédent)
```

```
graph_gte3 = ggplot(data, aes(x=Wages, y=Pouvoir, colour=Siècle))+
  geom_point()+
  geom_smooth(se=FALSE, linewidth=0.5)+
  scale_x_continuous(name="Salaire hebdomadaire (shillings)", breaks=seq(0,30,5))+
  scale_y_continuous(name="Pouvoir d'achat (kg de blé)", breaks=seq(0,4,0.5))+
  scale_color_manual(values=c("blue4", "coral", "#5ab4ac", "darkmagenta"))+
  theme_bw()+
  labs(colour="Siècle")+
  theme(legend.background=element_rect(fill="grey81"), legend.key.size=unit(1, "cm"))+
  ggtitle("Pouvoir d'achat en fonction du salaire")
graph_gte3
```



Discussion

La représentation des données me semble plus claire sur le premier graphique de cette section, où l'on peut voir l'évolution du pouvoir d'achat en fonction du temps. En effet, je trouve que cela montre bien que le pouvoir d'achat des ouvriers a augmenté au cours du temps, même si néanmoins à partir du 19^e siècle il semble avoir tendance à diminuer de nouveau. Les "barres" permettent également d'observer les fluctuations de ce pouvoir d'achat.