

Le pouvoir d'achat des ouvriers anglais du XVIe au XIXe siècle

Antoine RICHARD

May 1, 2020

Contents

1	Préparation des données	1
1.1	Téléchargement des données	2
1.2	Recherches des données manquantes	3
1.3	Extractions des colonnes utilisées	3
1.4	Vérification des données	4
1.5	Vérifications des dates	4
2	Reproduction des résultats de William Playfair	5
2.1	Évolution du prix du blé	5
2.2	Évolution des salaires	6
2.3	Ajout de divers éléments	7
3	Mise à jour du graphe de Playfair	8
3.1	Différenciation entre Prix et Salaires	8
3.2	Autres représentations graphiques	9
4	Mise en évidence du pouvoir d'achat des ouvriers	12
4.1	Évolution des salaires en fonction du prix du blé	14

Dans document computanionel nous cherchons à reproduire et mettre à jour l'étude de William Playfair sur l'évolution du pouvoir d'achat des ouvriers entre les année 1565 et 1810.

1 Préparation des données

Les données de l'étude de William Playfair sont disponible à l'adresse suivante: <https://raw.githubusercontent.com/vincentarelbundock/Rdatasets/>

```
master/csv/HistData/Wheat.csv
```

Quelques explications des colonnes données sur le site du MOOC:

Nom de colonne	Libellé de colonne
	Numéro de la ligne
Year	Année au format YYYY
Wheat	Le prix en shillings pour un quart de boisseau de blé
Wages	Salaire en shillings par semaine

Notons que, jusqu'en 1971, la livre sterling était divisée en 20 shillings, et un shilling en 12 pences.

Notons aussi qu'un quart de boisseau équivaut 15 livres britanniques ou 6,8 kg.

1.1 Téléchargement des données

Afin de ne pas avoir à télécharger les données à chaque execution, et pour pouvoir travailler dessus sans accès internet, nous sauvegardons ces données dans un fichier csv.

De plus, afin d'éviter de possible problèmes d'encodage, nous enregistrons le fichier octet par octet.

```
from urllib.request import urlopen
from os.path import exists

data_file = "data.csv"
if not exists(data_file):
    f = open(data_file,"wb")
    f.write(urlopen(data_url).read())
    f.close()
```

Après avoir téléchargé les données (si nécessaire), nous commençons par charger les données qui nous intéressent dans un tableau.

Regardons un résumé de nos données:

```
summary(table)

Min. : 1      Min. :1565    Min. :26.00   Min. : 5.000
1st Qu.:14    1st Qu.:1630  1st Qu.:33.00 1st Qu.: 6.145
Median :27    Median :1695  Median :41.00 Median : 7.800
Mean :27     Mean :1695   Mean :43.26   Mean :11.582
3rd Qu.:40   3rd Qu.:1760 3rd Qu.:47.00 3rd Qu.:14.875
Max. :53     Max. :1821   Max. :99.00   Max. :30.000
nil         nil         nil          NA's :3
```

Nous pouvons déjà voir que certaines données semblent manquer: trois entrées n'ont pas de valeur pour la colonne "Wages".

1.2 Recherches des données manquantes

Afin de ne pas gêner de futurs traitements de données, commençons par détecter et supprimer les entrées sans donnée pour la colonne "Wages".

```
table = table[!is.na(table$Wages),]  
summary(table)
```

X	Year	Wheat	Wages
Min. : 1.00	Min. :1565	Min. :26.00	Min. : 5.000
1st Qu.:13.25	1st Qu.:1626	1st Qu.:32.25	1st Qu.: 6.145
Median :25.50	Median :1688	Median :40.25	Median : 7.800
Mean :25.50	Mean :1688	Mean :42.14	Mean :11.582
3rd Qu.:37.75	3rd Qu.:1749	3rd Qu.:45.75	3rd Qu.:14.875
Max. :50.00	Max. :1810	Max. :99.00	Max. :30.000

1.3 Extractions des colonnes utilisées

Nous pouvons remarquer que la première colonne du tableau ne contient que les numéros des lignes. Dans notre étude seuls les colonnes Year, Wheat et Wages nous intéressent. Nous supprimons donc la première colonne de nos données.

```
table = table[,2:4]  
summary(table)
```

Year	Wheat	Wages
Min. :1565	Min. :26.00	Min. : 5.000
1st Qu.:1626	1st Qu.:32.25	1st Qu.: 6.145
Median :1688	Median :40.25	Median : 7.800
Mean :1688	Mean :42.14	Mean :11.582
3rd Qu.:1749	3rd Qu.:45.75	3rd Qu.:14.875
Max. :1810	Max. :99.00	Max. :30.000

1.4 Vérification des données

Vérifions maintenant si les données de notre tableau sont crédibles selon les critères suivants:

- Year: nombre entier composé de quatre chiffres
- Wheat et Wages: valeur numérique

```
for(row in 1:nrow(table)){
  if(!is.integer(table[row,"Year"])
     || nchar(table[row,"Year"]) != 4)
  {
    print("Valeur suspecte dans la colonne 'Year': ")
    print(table[row,])
  }
  if(!is.numeric(table[row,"Wheat"])){
    print("Valeur suspecte dans la colonne 'Wheat': ")
    print(table[row,])
  }
  if(!is.numeric(table[row,"Wages"])){
    print("Valeur suspecte dans la colonne 'Wages': ")
    print(table[row,])
  }
}
```

Tout semble en règles.

1.5 Vérifications des dates

Chaque entrée est supposée être séparé d'exactly cinq années, vérifions cela.

```
for(row in 2:nrow(table)){
  if(table[row,"Year"] - table[row-1,"Year"] != 5){
    print(
      paste(
        "Il y a",
        table[row,"Year"] - table[row-1,"Year"],
        "ans entre", table[row,"Year"],
        "et", table[row-1,"Year"]
      )
    )
  }
}
```

```
    )  
  }  
}
```

Pas de problème de ce côté là non plus, nous pouvons passer à l'analyse de nos données.

2 Reproduction des résultats de William Playfair

Avant d'effectuer nos propres opérations sur les données recueillies, tentons de reproduire le graphe proposé par William Playfair dans son étude.

Celui-ci représentait l'évolution du prix du blé à l'aide de barres, un barre par tranche de 5 années, et l'évolution des salaires à l'aide d'une aire bleue délimitée d'une courbe rouge.

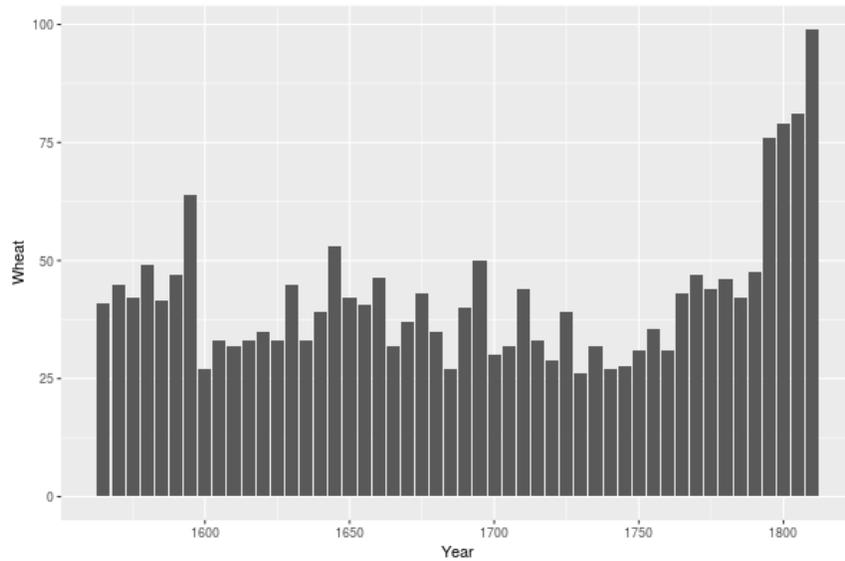
Pour reproduire ce graphe nous allons utilisé la librairie *ggplot2*, qui permet une création de graphe simple, étape par étape.

```
library(ggplot2)
```

2.1 Évolution du prix du blé

Comme introduit plus tôt, nous souhaitons visualiser l'évolution du prix du blé au cours du temps à l'aide d'un *barplot*, que nous obtenons via la fonction *geom_bar()*.

```
p <- ggplot(table, aes(x=Year)) +  
  geom_bar(stat="identity", aes(y=Wheat))  
p
```



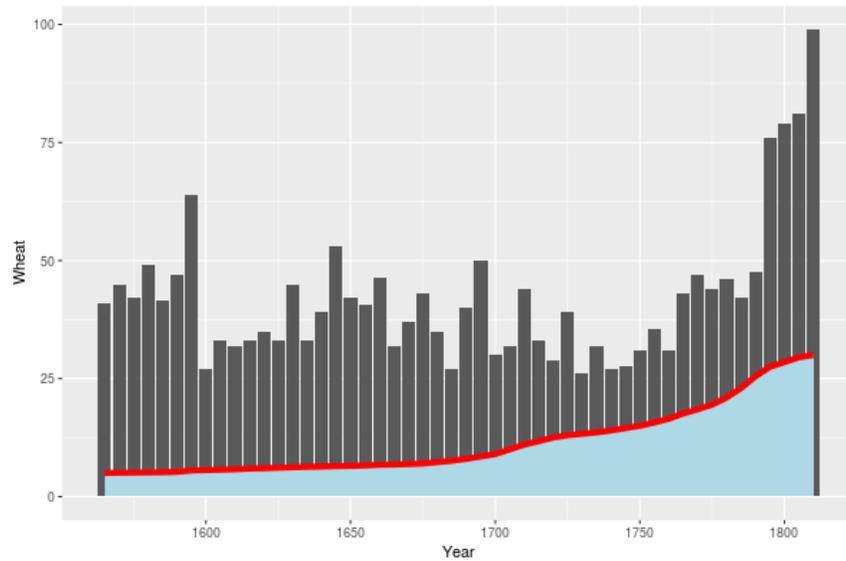
Notons que, puisque nous avons supprimée les entrées pour lesquelles nous n'avions pas de valeurs pour la colonne *Wages*, les trois dernières barre du graphe original ne sont pas présentes dans notre reproduction.

2.2 Évolution des salaires

Dans l'étude de Playfair, l'évolution des salaires était représentée à l'aide d'une aire bleue délimitée par une courbe rouge.

Pour reproduire cela nous utilisons la fonction *geom_area()* pour l'aire et la fonction *geom_line()* pour la ligne de délimitation.

```
p <- p + geom_area(aes(y=Wages),stat="identity",fill="lightblue")
p <- p + geom_line(aes(y=Wages),color="red",size=2)
p
```



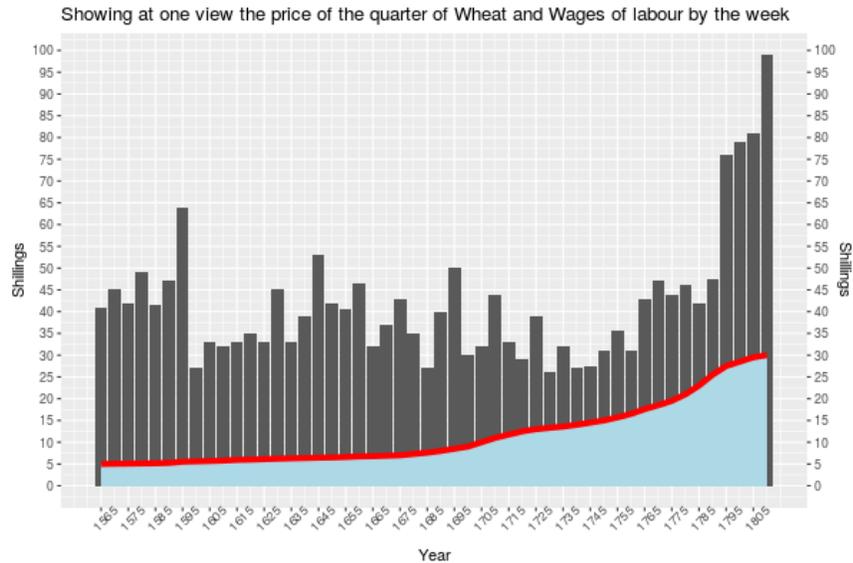
2.3 Ajout de divers éléments

Maintenant que nous avons affiché nos données dans un graphe à la manière de Playfair, nous allons procéder à quelques ajouts pour nous rapprocher du graphe initial, tels que:

- Changer le label de l'axe des ordonnées en "Shillings"
- Dupliquer l'axe des ordonnées à droite du graphe
- Répartir l'axe des ordonnées en divisions de 5 shillings
- Répartir l'axe des abscisses en divisions de 10 années
- Ajouter un titre

```
p <- p + ylab("Shillings")
p <- p + scale_y_continuous(
  breaks=seq(0,100,5),
  sec.axis=sec_axis(~.,name=derive(),breaks=derive())
)
p <- p + scale_x_continuous(
  breaks=seq(min(table$Year),max(table$Year),10)
) +
  theme(axis.text.x = element_text(angle=45))
```

```
p <- p + ggtitle("Showing at one view the price of the quarter of Wheat and Wages of labour by the week")
p
```



3 Mise à jour du graphe de Playfair

Lors de la sortie de ses travaux, Playfair était un pionnier de la présentation graphique des données.

Cependant, certains choix de représentation ne sont plus admissibles de nos jours.

Dans cette section nous proposons des modifications pour mettre le graphe de Playfair en adéquation avec les normes actuelles.

3.1 Différenciation entre Prix et Salaires

Dans son graphe, Playfair combine "Shillings par quart de boisseau de blé" et "Shillings par semaine" en une seule unité "Shillings", ce qui n'est plus acceptable de nos jours.

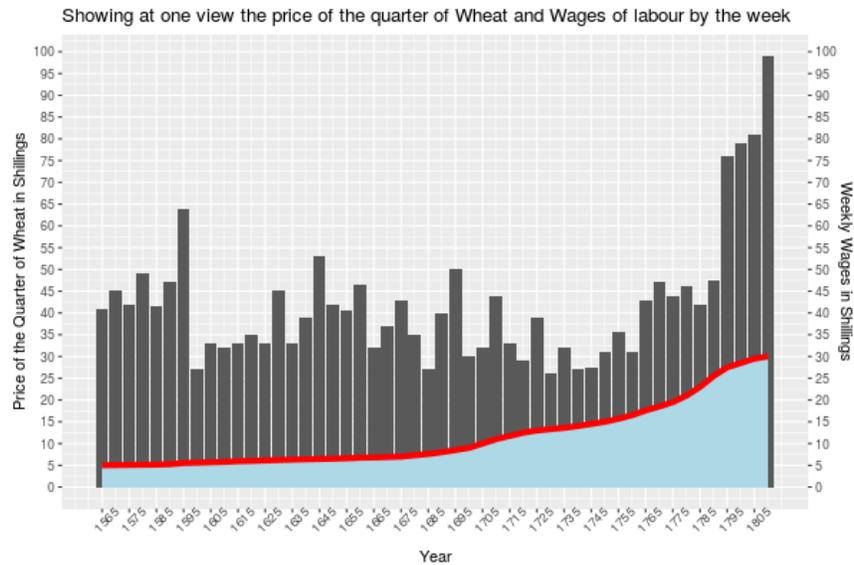
Notre première modification du graphe de Playfair consistera donc à différencier les unités "Shillings par quart de boisseau de blé" et "Shillings par semaine" en utilisant deux axes distincts.

Pour ce faire nous utilisons simplement le paramètre *sec.axis* de la fonction *scale_y_continuous()*.

```

ggplot(table, aes(x=Year)) +
  geom_bar(aes(y=Wheat), stat="identity") +
  geom_area(aes(y=Wages), stat="identity", fill="lightblue") +
  geom_line(aes(y=Wages), stat="identity", color="red", size=2) +
  ylab("Price of the Quarter of Wheat in Shillings") +
  scale_y_continuous(
    breaks=seq(0,100,5),
    sec.axis=sec_axis(~.,
                      name="Weekly Wages in Shillings",
                      breaks=derive()
                    )
  )+
  scale_x_continuous(breaks=seq(min(table$Year),max(table$Year),10))+
  theme(axis.text.x = element_text(angle=45)) +
  ggtitle("Showing at one view the price of the quarter of Wheat and Wages of labour

```



3.2 Autres représentations graphiques

Mixer ensemble un *barplot* et un *linechart*, mis à part dans des cas très spécifiques, peut rendre compliquer la comparaison des données.

De nos jours, il est préférable d'utiliser le même type de représentation graphique lors de la comparaison entre deux données.

Cependant, il nous faut dans un premier temps retravailler le format de

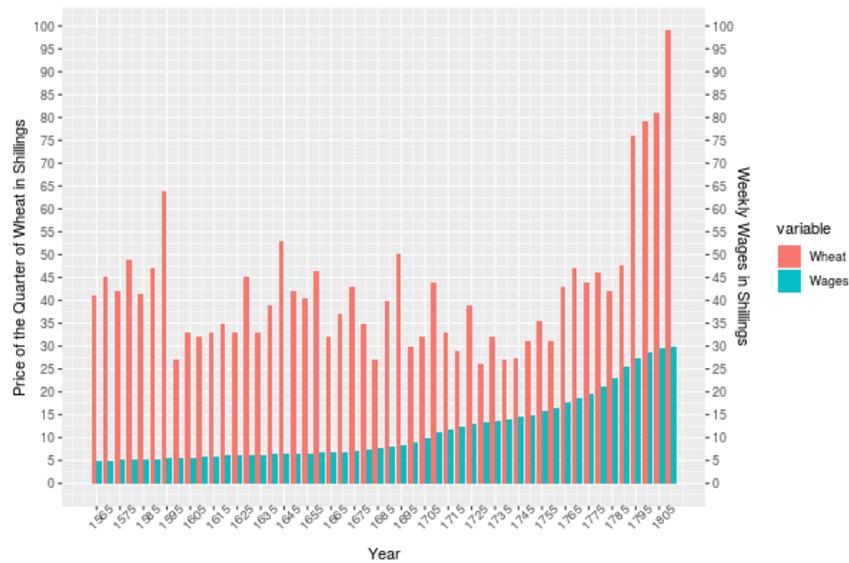
nos données à l'aide de la fonction `melt()` de la librairie `reshape2`, afin de rendre plus aisée leur représentation à l'aide de `ggplot`.

```
library(reshape2)
n_data <- melt(table, id.vars="Year", measure.vars=c("Wheat","Wages"))
summary(n_data)
```

	Year	variable	value
Min.	:1565	Wheat:50	Min. : 5.00
1st Qu.	:1625	Wages:50	1st Qu.: 7.90
Median	:1688		Median :27.25
Mean	:1688		Mean :26.86
3rd Qu.	:1750		3rd Qu.:40.12
Max.	:1810		Max. :99.00

Une première représentation possible est d'utiliser des barres les prix et pour les salaires, cela afin de rendre plus aisée leur comparaison.

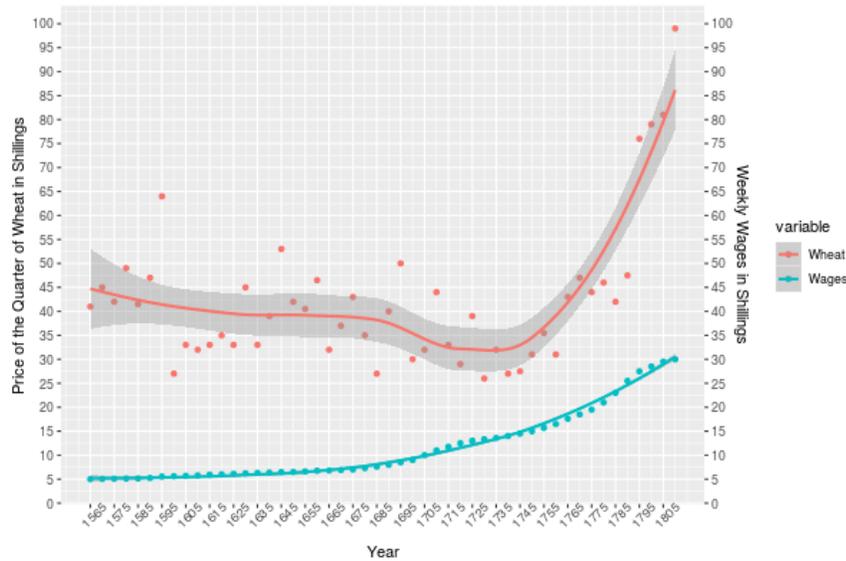
```
ggplot(n_data, aes(x=Year, y=value, fill=variable)) +
  geom_bar(stat="identity", position=position_dodge()) +
  ylab("Price of the Quarter of Wheat in Shillings") +
  scale_y_continuous(
    breaks=seq(0,100,5),
    sec.axis=sec_axis(~.,
                      name="Weekly Wages in Shillings",
                      breaks=derive()
                    )
  )+
  scale_x_continuous(breaks=seq(min(table$Year),max(table$Year),10)) +
  theme(axis.text.x = element_text(angle=45))
```



Cependant, si cette représentation permet de comparer prix du blé et salaire hebdomadaire, la prix du blé reste assez volatile et il est compliqué d’observer l’évolution conjointe de ces deux données.

Pour ce faire nous allons plutôt utiliser une représentation par nuage de points et régressions à l’aide des fonctions `geom_point()` et `geom_smooth()`.

```
ggplot(n_data,aes(x=Year,y=value, color=variable)) +
  geom_point() +
  geom_smooth() +
  ylab("Price of the Quarter of Wheat in Shillings") +
  scale_y_continuous(
    breaks=seq(0,100,5),
    sec.axis=sec_axis(~.,
                      name="Weekly Wages in Shillings",
                      breaks=derive()
                    )
  )+
  scale_x_continuous(breaks=seq(min(table$Year),max(table$Year),10)) +
  theme(axis.text.x = element_text(angle=45))
```



Avec cette dernière représentation, nous pouvons observer une augmentation quasi-linéaire des salaires depuis environ l'année 1700, ainsi qu'une diminution du prix du blé jusqu'à environ l'année 1750, ce qui peut laisser supposer à une augmentation du "pouvoir d'achat" des ouvriers.

Cependant, nous pouvons aussi observer qu'à partir de 1750 (environ), le prix du blé a augmenté de manière drastiquement plus rapidement que les salaires hebdomadaires.

Cette observation remet quelque peu en question l'augmentation du pouvoir d'achat des ouvriers décrite par William Playfair.

4 Mise en évidence du pouvoir d'achat des ouvriers

Dans son étude, William Playfair souhaitait mettre en évidence l'augmentation dans le temps du pouvoir d'achat des ouvriers.

Cependant, le pouvoir d'achat, définie comme "la quantité de blé qu'un ouvrier peut acheter avec son salaire hebdomadaire", n'est pas une quelque chose que l'on peut quantifier simplement en comparant l'évolution du prix du blé et des salaires. Il nous faut donc la calculer pour en avoir une meilleure appréciation.

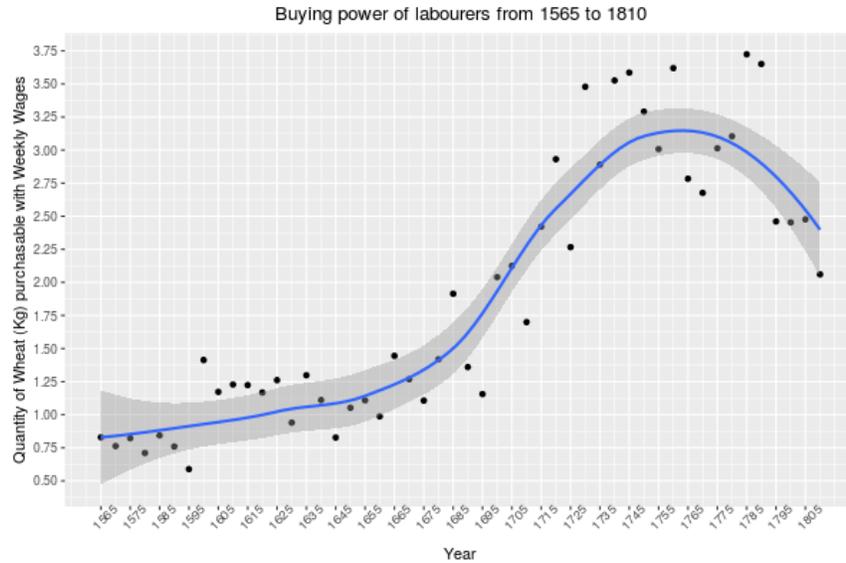
Pour ce faire, nous allons ajouter une colonne Qty à notre table qui contiendra la quantité de blé (en Kg) qu'un ouvrier peut acheter avec son salaire hebdomadaire, calculé à l'aide de la formule $Qty = Wages/Wheat * 6.8$ (un quart de boisseau de blé équivalent, environ, à 6.8Kg).

```
table["Qty"] = table$Wages / table$Wheat * 6.8
summary(table)
```

Year	Wheat	Wages	Qty
Min. :1565	Min. :26.00	Min. : 5.000	Min. :0.5886
1st Qu.:1626	1st Qu.:32.25	1st Qu.: 6.145	1st Qu.:1.1088
Median :1688	Median :40.25	Median : 7.800	Median :1.4316
Mean :1688	Mean :42.14	Mean :11.582	Mean :1.9012
3rd Qu.:1749	3rd Qu.:45.75	3rd Qu.:14.875	3rd Qu.:2.7566
Max. :1810	Max. :99.00	Max. :30.000	Max. :3.7238

Nous pouvons maintenant représenter l'évolution du pouvoir d'achat des ouvriers au court du temps comme nous l'avons fait précédemment à l'aide d'un nuage de points et d'un regression.

```
ggplot(table, aes(x=Year, y=Qty)) +
  geom_point() +
  geom_smooth() +
  scale_y_continuous(breaks=seq(0,5,0.25)) +
  scale_x_continuous(breaks=seq(min(table$Year),max(table$Year),10)) +
  ylab("Quantity of Wheat (Kg) purchasable with Weekly Wages") +
  ggtitle("Buying power of labourers from 1565 to 1810") +
  theme(
    axis.text.x = element_text(angle=45),
    plot.title = element_text(hjust = 0.5)
  )
```



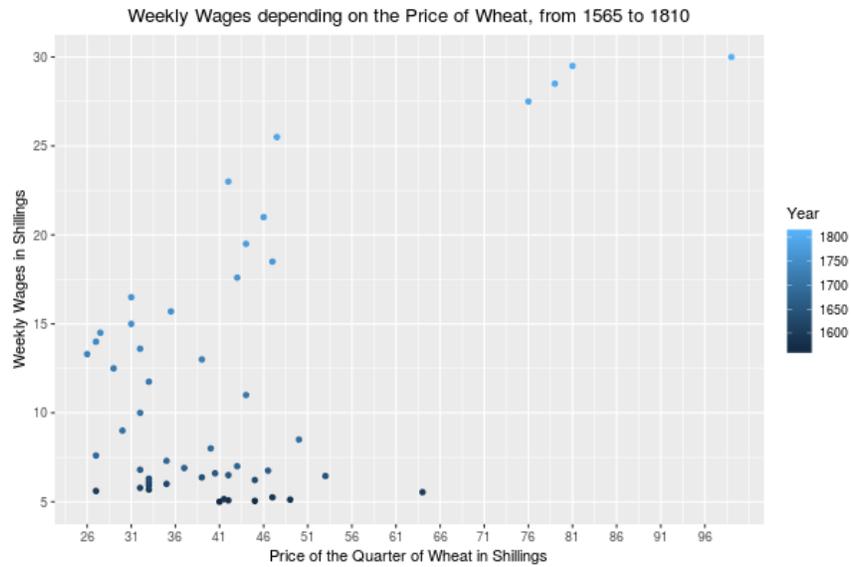
Nous pouvons voir sur ce graphique une nette augmentation du pouvoir d'achat des ouvriers avec un pic aux alentours des années 1750 - 1770. Cependant, nous pouvons aussi observer une diminution du pouvoir d'achat à partir de ces années, ce qui peut s'expliquer par l'augmentation drastique du prix du blé depuis 1750, comme observé précédemment.

4.1 Évolution des salaires en fonction du prix du blé

Une autre manière de représenter le pouvoir d'achat, peut se faire en observant l'évolution du salaire par rapport au prix du blé à l'aide d'un nuage de points.

Afin de ne pas perdre la progression du temps, nous colorons les points en fonction de l'année associée. Les points les plus sombres représentant les plus vieilles entrées et les points les plus clairs, les plus récentes.

```
ggplot(table, aes(x=Wheat,y=Wages)) +
  geom_point(aes(colour=Year)) +
  ylab("Weekly Wages in Shillings") +
  xlab("Price of the Quarter of Wheat in Shillings") +
  scale_y_continuous(breaks=seq(min(table$Wages),max(table$Wages),5)) +
  scale_x_continuous(breaks=seq(min(table$Wheat),max(table$Wheat),5))+
  ggtitle("Weekly Wages depending on the Price of Wheat, from 1565 to 1810") +
  theme(plot.title = element_text(hjust = 0.5))
```



Nous pouvons voir sur ce graphique que, à quelques exceptions près, le prix du quart de boisseau de blé reste entre 26 et 51 Shillings au cours du temps. Au contraire, le salaire hebdomadaire des ouvriers n'a fait qu'augmenter au cours du temps, laissant supposer une augmentation du pouvoir d'achat des ouvriers.