

Peer-evaluated exercise

Purchasing power of English workers from the 16th to the 19th century

Yevheniya Nosyk (M2 MOSIG)

12 Dec 2020

The dataset

The goal of the exercise is to reproduce the study of William Playfair on the purchasing power of English workers. The scientist summarized his analysis on a graph (https://fr.wikipedia.org/wiki/William_Playfair#/media/File:Chart_Showing_at_One_View_the_Price_of_the_Quarter_of_Wheat,_and_Wages_of_Labour_by_the_Week,_from_1565_to_1821.png) but did not provide the raw data. We will load the dataset recreated later that represents the points on the original graph:

```
data = read.csv("wheat.txt", header = TRUE)
head(data)
```

```
##   X Year Wheat Wages
## 1 1 1565  41.0  5.00
## 2 2 1570  45.0  5.05
## 3 3 1575  42.0  5.08
## 4 4 1580  49.0  5.12
## 5 5 1585  41.5  5.15
## 6 6 1590  47.0  5.25
```

```
sprintf("There are %s observations and %s features.", nrow(data), ncol(data))
```

```
## [1] "There are 53 observations and 4 features."
```

Let's examine every column in more details. The column X does not hold any valuable information, except of numbering every row. Thus, we will remove it.

```
data = subset(data, select = -c(X))
```

We see that for each column the datatype was inferred correctly (integer for Year and double for Wheat/Wages). We can print the minimum and maximum values of every column to check that they actually make sense. For example, we know that observations took place between 1565 and 1821. We also do not expect negative values for Wheat and Wages:

```
sprintf("The minimum year is %s and the maximum %s", min(data$Year), max(data$Year))
```

```
## [1] "The minimum year is 1565 and the maximum 1821"
```

```
sprintf("The minimum amount of wheat is %s and the maximum %s",min(data$Wheat),
      max(data$Wheat))
```

```
## [1] "The minimum amount of wheat is 26 and the maximum 99"
```

```
sprintf("The minimum wage is %s and the maximum %s",min(data$Wages), max(data$Wages))
```

```
## [1] "The minimum wage is NA and the maximum NA"
```

We just discovered that some of the observations do not contain data for wages. Locate those:

```
subset(data, is.na(data$Wages))
```

```
##      Year Wheat Wages
## 51 1815      78    NA
## 52 1820      54    NA
## 53 1821      54    NA
```

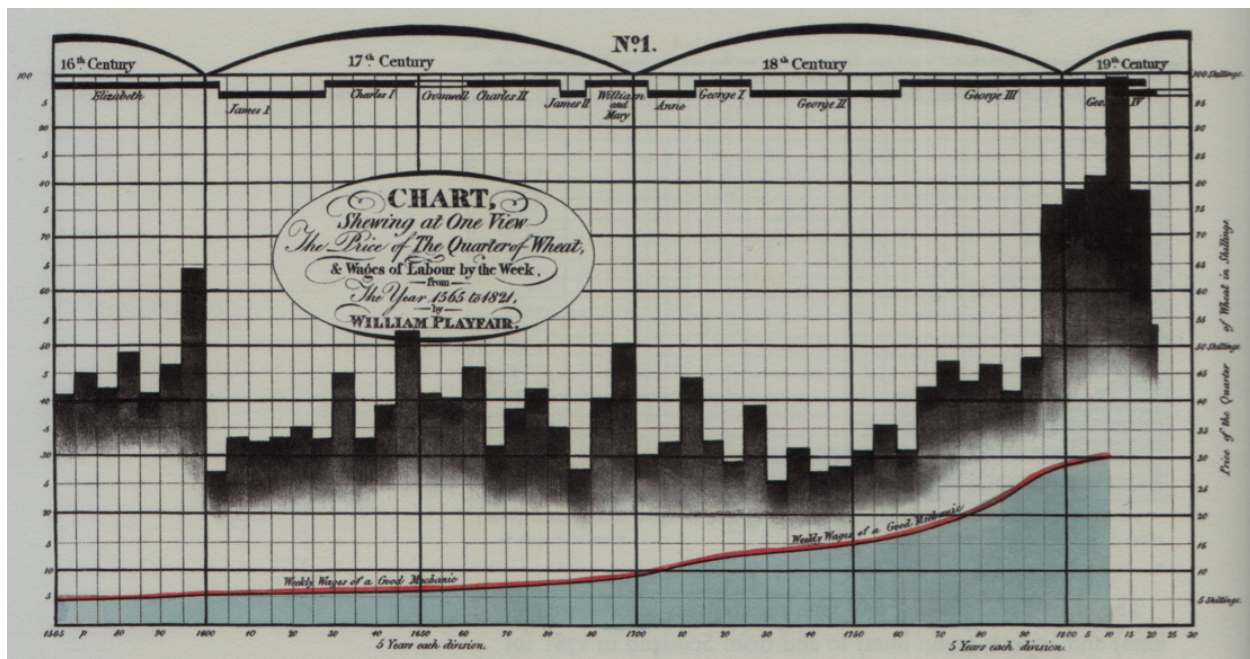
There are three rows with missing data. Looking at the original graph, those correspond to the last observations on the bar chart and wages are indeed missing. We would usually remove such observations, but we will not do so to reproduce the original graph exactly. We still want to check the values of Wages column, disregarding missing ones:

```
sprintf("The minimum wage is %s and the maximum %s",min(data$Wages,na.rm=TRUE),
      max(data$Wages,na.rm=TRUE))
```

```
## [1] "The minimum wage is 5 and the maximum 30"
```

Part 1: Recreate the original graph

Below is a graph drawn by William Playfair:



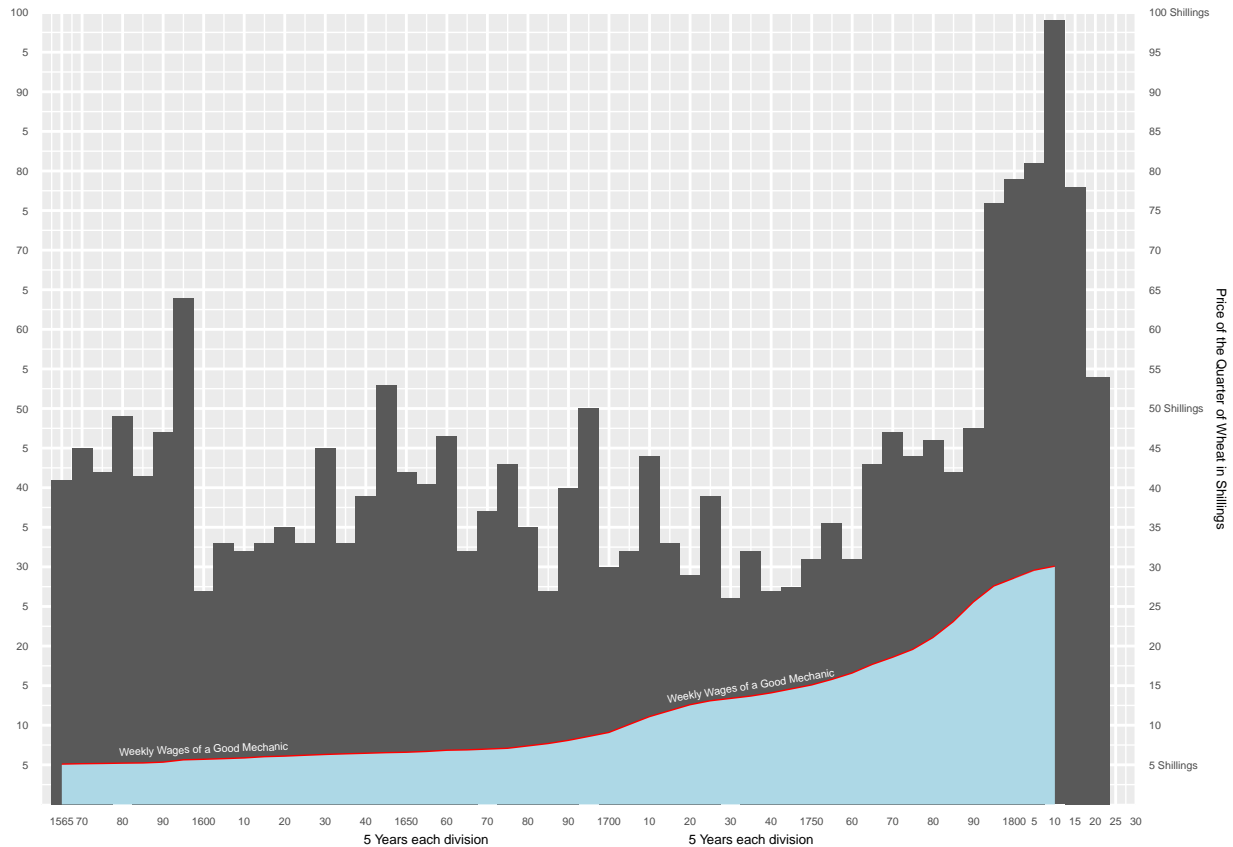
We will use ggplot to reproduce the following key points:

- X axis represents the years of observations, from 1565 to 1830. The axis tick labels are not homogeneous though. They are spaced out at different intervals - either 5 (at the very beginning and at the very end) or 10 years.
- Y axis represents the amount of shillings that measure the price of a quarter of a pound of wheat as well as the weekly salary.
- The red line illustrates the evolution of salaries in time.
- The bar chart illustrates the evolution of the price of a quarter of a pound of wheat.

```
library(ggplot2)

ggplot(data=data) +
  geom_bar(aes(Year,Wheat), stat='identity', width=5) +
  geom_line(aes(Year,Wages),color="red") +
  geom_area(aes(Year,Wages),fill = "lightblue") +
  scale_x_continuous(breaks=c(1565,1570,1580,1590,1600,1610,1620,1630,1640,1650,1660,
                              1670,1680,1690,1700,1710,1720,1730,1740,1750,1760,1770,
                              1780,1790,1800,1805,1810,1815,1820,1825,1830),
                    labels=c("1565", "70", "80", "90", "1600", "10", "20", "30", "40", "1650",
                              "60", "70", "80", "90", "1700", "10", "20", "30", "40", "1750",
                              "60", "70", "80", "90", "1800", "5", "10", "15", "20", "25", "30"),
                    expand = c(0, 0),limits = c(1560, 1830))+
  scale_y_continuous(breaks = c(5,10,15,20,25,30,35,40,45,50,55,
                                60,65,70,75,80,85,90,95,100),
                    labels=c("5 Shillings", "10", "15", "20", "25", "30", "35", "40", "45",
                              "50 Shillings", "55", "60", "65", "70", "75", "80", "85", "90",
                              "95", "100 Shillings"),
                    position = "right",
                    expand = c(0, 0), limits = c(0, 100),
                    sec.axis = sec_axis(~., breaks = c(5,10,15,20,25,30,35,40,45,50,55,
                                                        60,65,70,75,80,85,90,95,100),
                    labels=c("5", "10", "5", "20", "5", "30", "5",
                              "40", "5", "50", "5", "60", "5", "70",
                              "5", "80", "5", "90", "5", "100")))+

  theme(
    axis.title.x = element_text(size=5),
    axis.title.y = element_text(size=5),
    axis.ticks.x = element_blank(),
    axis.ticks.y = element_blank(),
    axis.text.x = element_text(size=4),
    axis.text.y = element_text(size=4),
  ) +
  annotate("text", x=1600, y=7, label= "Weekly Wages of a Good Mechanic",
          size=1.25, angle=2, color="white") +
  annotate("text", x=1735, y=15, label= "Weekly Wages of a Good Mechanic",
          size=1.25, angle=9, color="white") +
  labs(x=paste("5 Years each division",
               "5 Years each division"),
       y="Price of the Quarter of Wheat in Shillings")
```



Part 2: Improve the original graph

There are several ways in which we can improve the original graph:

- use both the right and left parts of the Y axis to represent two different quantities - “shillings per week” and “the price of the quarter of a pound”;
- improve the representation of the left Y axis by showing full numbers instead of “5”s;
- X axis will have its ticks spaced out evenly, at the interval of 10 years (except for the very beginning, to make years end with 0s for better readability);
- X axis will end at year 1810, as there are no more observations after;
- use line charts for both Wages and Wheat and add legend;
- remove the three data points that do not have the corresponding wage values;
- rename the X axis;
- add the graph title;

```
data <- subset(data, !is.na(data$Wages))

library(ggplot2)

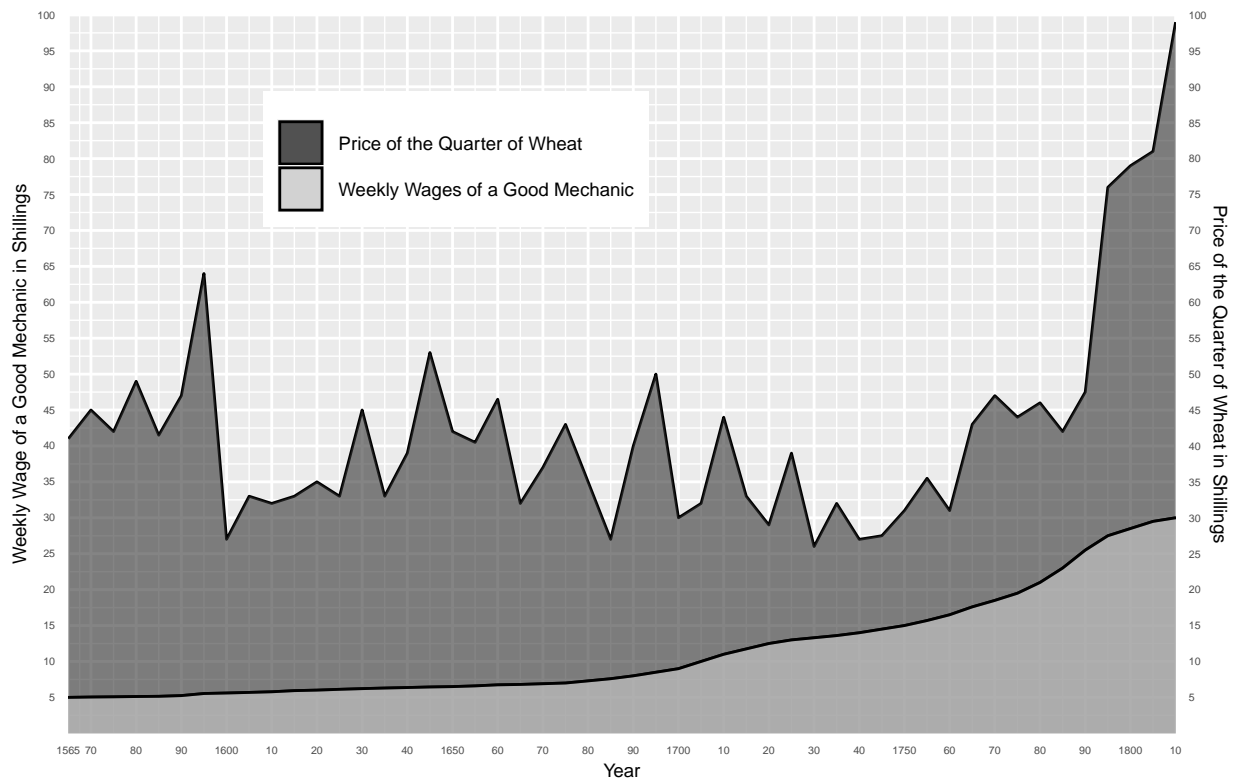
ggplot(data=data, aes(x=Year)) +
  geom_line(aes(y=Wheat)) + geom_line(aes(y=Wages)) +
  geom_area(aes(y=Wheat,fill="Price of the Quarter of Wheat"),alpha=0.6) +
  geom_area(aes(y=Wages,fill="Weekly Wages of a Good Mechanic"),
    colour = "black",alpha=0.6) +
```

```

scale_fill_grey() +
scale_x_continuous(breaks=c(1565,1570,1580,1590,1600,1610,1620,1630,1640,1650,1660,
                             1670,1680,1690,1700,1710,1720,1730,1740,1750,1760,1770,
                             1780,1790,1800,1810),
                  labels=c("1565", "70", "80", "90", "1600", "10", "20", "30", "40", "1650",
                             "60", "70", "80", "90", "1700", "10", "20", "30", "40", "1750",
                             "60", "70", "80", "90", "1800", "10"),
                  expand = c(0, 0), limits = c(1565, 1810)) +
scale_y_continuous(breaks = seq(5,100,5), position = "right",
                  expand = c(0, 0), limits = c(0, 100),
                  sec.axis = sec_axis(~., breaks = seq(5,100,5),
                  name = "Weekly Wage of a Good Mechanic in Shillings")) +
theme(
  axis.title.x = element_text(size=7), axis.title.y = element_text(size=7),
  axis.ticks.x = element_blank(), axis.ticks.y = element_blank(),
  axis.text.x = element_text(size=4), axis.text.y = element_text(size=4),
  legend.position=c(0.35, 0.8),
  legend.text = element_text(size=7),
  legend.title = element_blank(),
  plot.title = element_text(size=10,face="bold",hjust = 0.5)
) +
labs(title=paste("Chart, Showing at One View The Price of The Quarter",
                  "of Wheat \n and Wages of Labour by the Week"),
     x="Year",
     y="Price of the Quarter of Wheat in Shillings")

```

**Chart, Showing at One View The Price of The Quarter of Wheat
and Wages of Labour by the Week**



Part 3: Evaluate the purchasing power

The current version of the plot shows how wheat price and weekly salary evolved with time. These absolute numbers do not give good intuition about the relationship between two variables. One may think that higher salary allows workers buy more wheat, but is it really the case? We will see by plotting the purchasing power, that is the number of quarters of pound of wheat a typical worker can buy weekly.

We first introduce a new column to our data, namely PrPow, which is merely the number of quarters of Wheat that can be bought with weekly salary. We will round the result to two digits after comma:

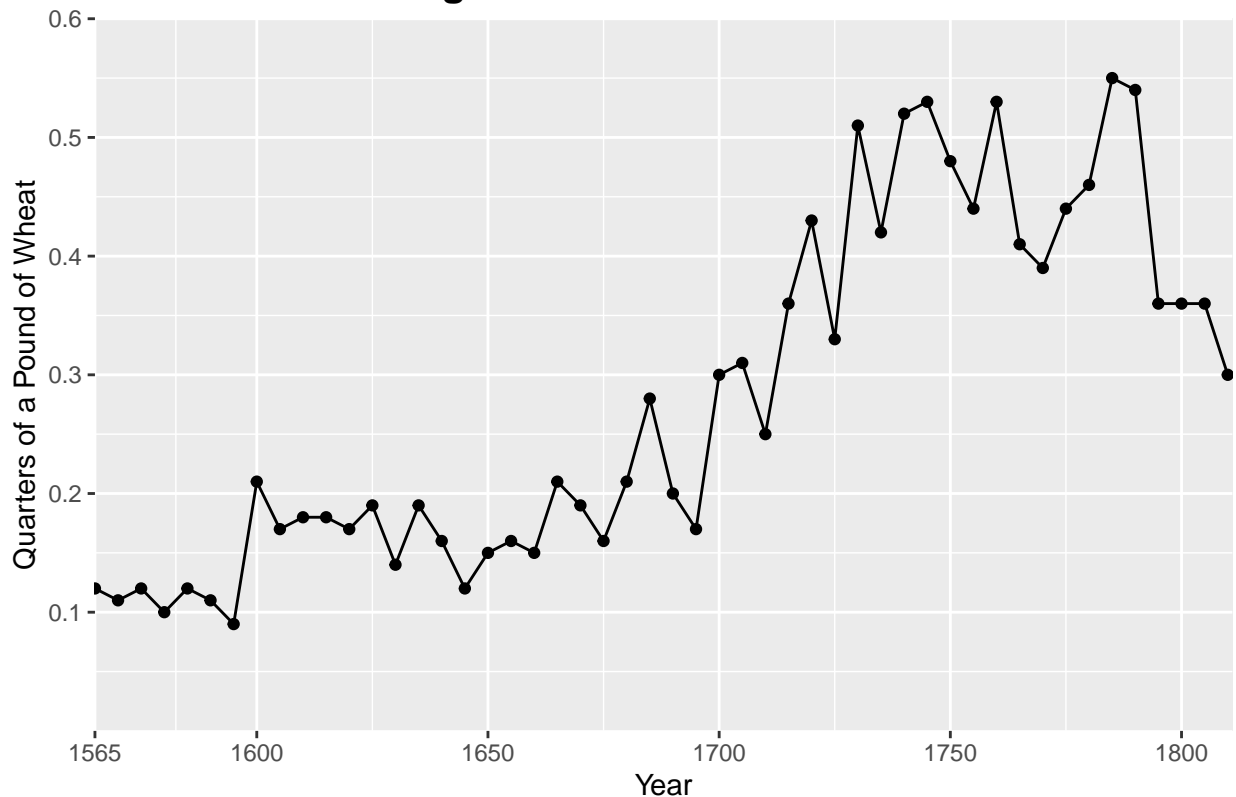
```
data$PrPow <- round(data$Wages/data$Wheat, digits = 2)
head(data)
```

```
##   Year Wheat Wages PrPow
## 1 1565  41.0   5.00  0.12
## 2 1570  45.0   5.05  0.11
## 3 1575  42.0   5.08  0.12
## 4 1580  49.0   5.12  0.10
## 5 1585  41.5   5.15  0.12
## 6 1590  47.0   5.25  0.11
```

We now plot the purchasing power as a function of time:

```
ggplot(data=data, aes(x=Year, y=PrPow)) +
  geom_line() +
  geom_point() +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 0.6),
                     breaks = seq(0.1, 0.6, 0.1)) +
  scale_x_continuous(breaks=c(1565, 1600, 1650, 1700, 1750, 1800),
                     expand = c(0, 0), limits = c(1565, 1811)) +
  theme(plot.title = element_text(size=15, face="bold", hjust = 0.5)) +
  labs(y="Quarters of a Pound of Wheat",
       title="The Purchasing Power of Workers from 1565 to 1810")
```

The Purchasing Power of Workers from 1565 to 1810



Even though we saw in the original graph that the salary has been constantly increasing, it did not always imply higher purchasing power. It is best seen during the last fifty years of observations - high salaries resulted in even higher wheat prices and, consequently, decreased purchasing power. This makes us doubt how well wages are correlated to wheat prices. We make a simple correlation test and pay a special attention to the correlation coefficient. It can be in the range $[-1,1]$, where -1 means strong negative correlation, $+1$ is strong positive correlation and 0 is no correlation at all. The obtained value is 0.58 , which signifies a moderate positive correlation between weekly wages and wheat prices. For 95% confidence interval, we want p-value to be less than 0.05 and it is indeed the case, meaning that the experiment results have a very low probability of being random.

```
cor.test(data$Wages, data$Wheat, method = "pearson")
```

```
##
## Pearson's product-moment correlation
##
## data: data$Wages and data$Wheat
## t = 4.9389, df = 48, p-value = 9.922e-06
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## 0.3603532 0.7393626
## sample estimates:
## cor
## 0.580477
```

We can plot the wages as a function of wheat price and add a regression line. To add the notion of Years, we will add labels to some of the data points:

```

library("ggpubr")
library("ggrepel")
ggplot(data, aes(x=Wheat, y=Wages, label=data$Year)) +
  geom_point() +
  geom_smooth(method=lm, level=0.95) +
  theme_minimal() +
  scale_y_continuous(expand = c(0, 0), limits = c(0, 34), breaks = seq(10,30,10)) +
  scale_x_continuous(expand = c(0, 0), limits = c(0, 105)) +
  labs(title = "Weekly Wages as a Function of the Price of Wheat") +
  theme(plot.title = element_text(size=15,face="bold",hjust = 0.5)) +
  geom_label_repel(
    aes(label=ifelse(as.logical(data$Year%%50),'',as.character(data$Year))),
    box.padding = unit(2.9, "lines"))

```

'geom_smooth()' using formula 'y ~ x'

