

# Mini Oktoberfest Analysis

Ranji Raj

2021-04-09

## Contents

Packages and data . . . . .	1
Beer consumption over time . . . . .	2
Price vs. consumption of Hendl over time . . . . .	3
Visitor totals before and after 9/11 . . . . .	5

In this mini analysis we work with a dataset provided by the Statistics Office Munich about the Oktoberfest. More specifically, the dataset includes annual Oktoberfest duration statistics, total visitors, daily visitors, beer prices, beer consumption, grilled chicken ( “*Hendl*”) prices and grilled chicken consumption.

Your task is to **fill in the blanks!**

## Packages and data

```
library(tidyverse)
of <- read_csv("oktoberfest-en.csv")
of
```

```
## # A tibble: 35 x 8
##   year duration_in_days visitors_total visitors_daily beer_price_per_l
##   <dbl>           <dbl>           <dbl>           <dbl>           <dbl>
## 1 1985             16             7.1             444             3.2
## 2 1986             16             6.7             419             3.3
## 3 1987             16             6.5             406             3.37
## 4 1988             16             5.7             356             3.45
## 5 1989             16             6.2             388             3.6
## 6 1990             16             6.7             419             3.77
## 7 1991             16             6.4             400             4.21
## 8 1992             16             5.9             369             4.42
## 9 1993             16             6.5             406             4.71
## 10 1994            16             6.6             413             4.89
## # ... with 25 more rows, and 3 more variables: beer_consumption_hl <dbl>,
## #   hendl_price <dbl>, hendl_consumption <dbl>
```

The dataset contains information on 35 years of the Oktoberfest. For each year, 7 statistics are available. Let's find out for which years exactly data are available.

```
of %>% pull(year)
```

```
## [1] 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999
## [16] 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014
## [31] 2015 2016 2017 2018 2019
```

In this analysis we'll focus on the change of the following measurements over time:

- `beer_price_per_l`: the beer price per liter in Euro
- `beer_consumption_hl`: the beer consumption in hectoliter (1 hl = 100 l)
- `hendl_price`: the price of a "Hendl" (grilled chicken) in Euro
- `hendl_consumption`: the number of Hendls eaten during the Oktoberfest

## Beer consumption over time

Which are the years with the highest beer consumption?

```
of %>%
  arrange(desc(year)) %>%
  select(year, beer_consumption_hl)
```

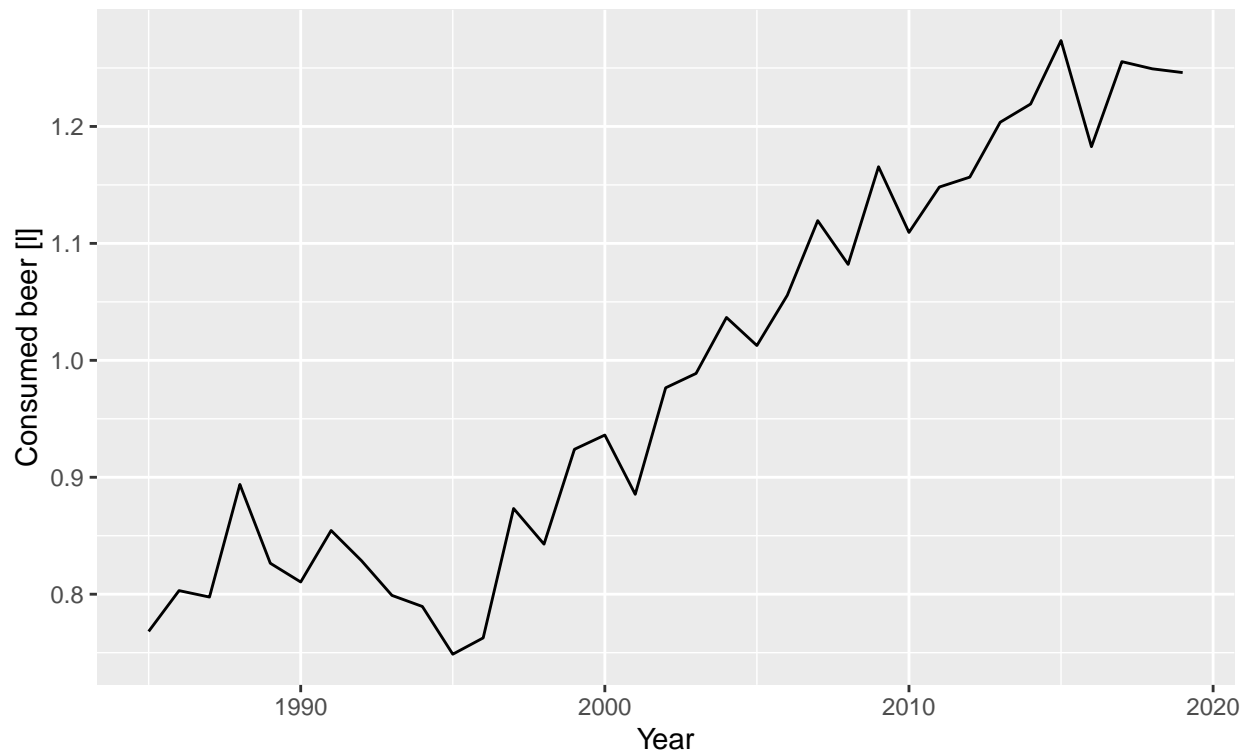
It can be concluded that there has been a higher consumption of beer in recent years.

A line plot gives us an even better overview of beer consumption over time:

```
of %>%
  mutate(beer_per_visitor_l = (100 * beer_consumption_hl) /
        (visitors_total * 1000000)) %>%
  ggplot(data = ., mapping = aes(x = year, y = beer_per_visitor_l)) +
  geom_line() +
  labs(
    title = "Average liters of beer consumed per visitor at the Oktoberfest",
    subtitle = "Time period: 1985 - 2019",
    x = "Year",
    y = "Consumed beer [l]"
  )
```

## Average liters of beer consumed per visitor at the Oktoberfest

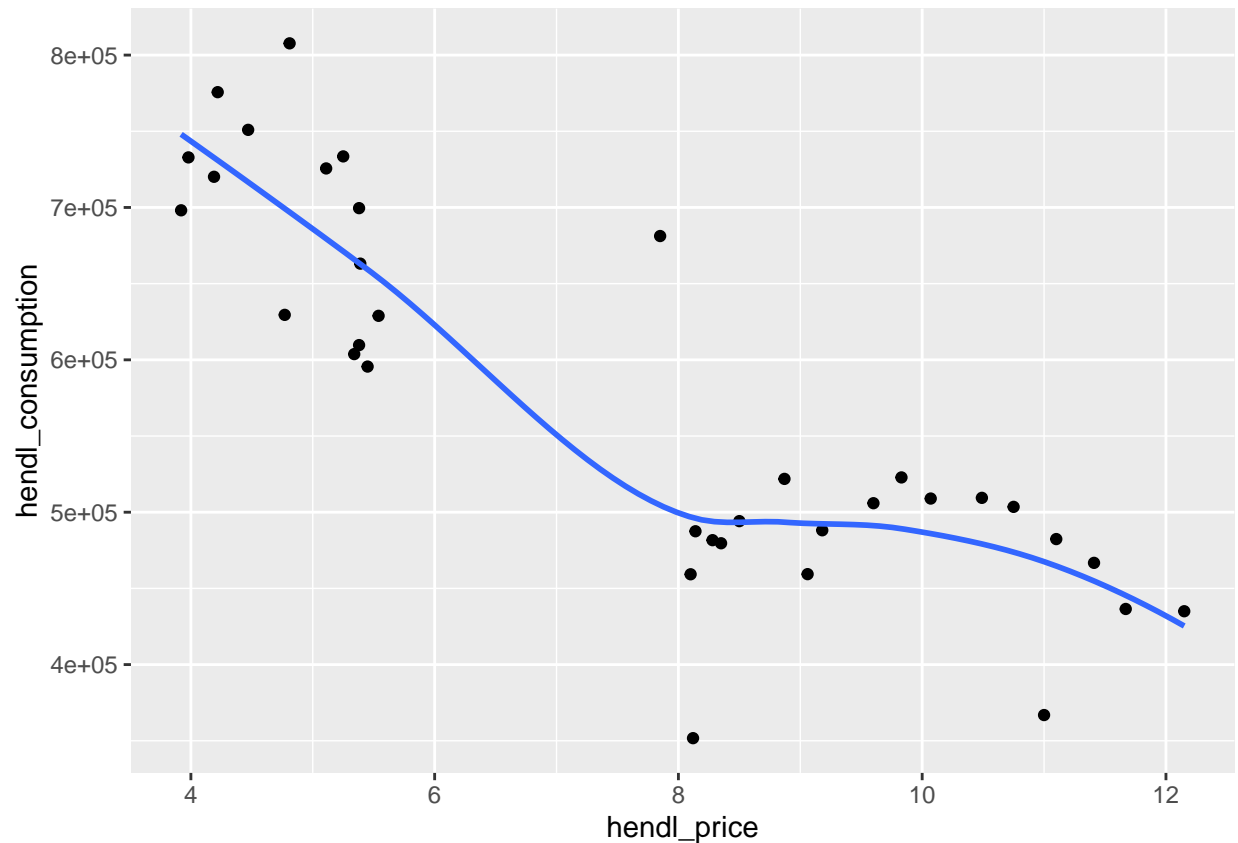
Time period: 1985 – 2019



## Price vs. consumption of Hendls over time

Next, let's make a plot to investigate whether there is a relationship between price and consumption of grilled chicken.

```
ggplot(data = of, mapping = aes(x = hendl_price, y = hendl_consumption)) +  
  geom_point() +  
  geom_smooth(se = FALSE)
```



What does this plot show us?

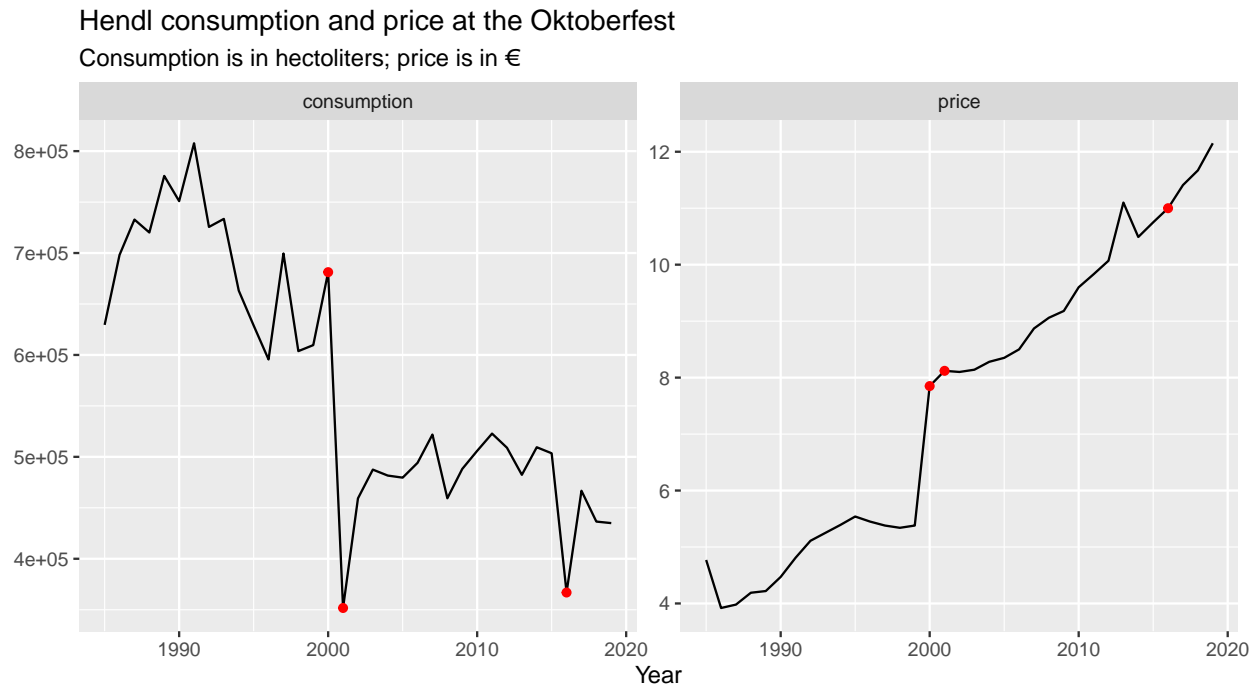
- In years with higher Hendl prices less Hendls are consumed.
- There are no points for prices between 6 and 7.5 euros. Was there perhaps a sharp increase in prices from one year to the next?
- There are three points that are far away from the overall trend.

Let us plot the change of Hendl price and consumption over time separately and highlight the “outlier years”.

```
ofp <- of %>%
  select(year, starts_with("hendl")) %>%
  mutate(is_outlier = case_when(
    hendl_price > 6 & hendl_consumption > 6e5 ~ TRUE,
    hendl_consumption < 4e5 ~ TRUE,
    TRUE ~ FALSE
  )) %>%
  pivot_longer(
    cols = starts_with("hendl"),
    names_to = "variable",
    values_to = "value",
    names_pattern = "hendl_(*)"
  )

ggplot(data = ofp, aes(x = year, y = value)) +
  geom_line() +
  geom_point(data = filter(ofp, is_outlier), color = "red") +
```

```
facet_wrap(~ variable, scales = "free_y") +
labs(
  x = "Year",
  y = NULL,
  title = "Hendl consumption and price at the Oktoberfest",
  subtitle = "Consumption is in hectoliters; price is in €"
)
```



The plot shows that there is a big drop in Hendl consumption in 2001. The Oktoberfest took place that year just days after Nine Eleven. Apparently, many people preferred to stay home this year due to safety concerns. This should reflect in the total number of visitors in that year.

## Visitor totals before and after 9/11

Create a line plot showing yearly visitor totals across years.

```
ggplot(data = of, aes(x = year, y = visitors_total)) +
  geom_line() +
  geom_smooth(se = FALSE) +
  geom_vline(xintercept = 2001, linetype = "dashed", color = "darkorange1", size = 1.5) +
  labs(
    title = "Yearly number of visitors at the Oktoberfest",
    x = "Year",
    y = "Visitors total [in millions]"
  )
```

Create a boxplot to show the distribution of visitor totals *before* and *after* 9/11.

```

ofm <- of %>%
  mutate(group = fct_inorder(ifelse(year >= 2011, "after 9/11", "before 9/11")))

ggplot(data = ofm, aes(x = year, y = visitors_total, group=1)) +
  geom_boxplot() +
  labs(
    title = "Visitors totals at the Oktoberfest before and after 9/11",
    x = NULL,
    y = "Visitors total [in millions]"
  )

```