

# Summary statistics + controls

Laura Menicacci

2023-02-26

## Contents

<b>Libraries &amp; data</b>	<b>1</b>
<b>Tidy data</b>	<b>2</b>
Add 0 values for missing years . . . . .	2
Filter for specific time frame . . . . .	2
<b>Summary statistics</b>	<b>2</b>
Boxplot - by year . . . . .	3
Boxplot - by US State . . . . .	3
<b>Add controls</b>	<b>4</b>
GDP - 1999-2020 . . . . .	4
Population - 1999-2020 . . . . .	4
R&D - 1999-2019 . . . . .	5
Crude oil prices 1999-2020 . . . . .	5
HDD & CDD . . . . .	6

## Libraries & data

```
## -- Attaching packages ----- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## Rows: 2288 Columns: 3
## -- Column specification -----
## Delimiter: ","
## chr (1): region
## dbl (2): prio_year, patent_count
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

## Tidy data

### Add 0 values for missing years

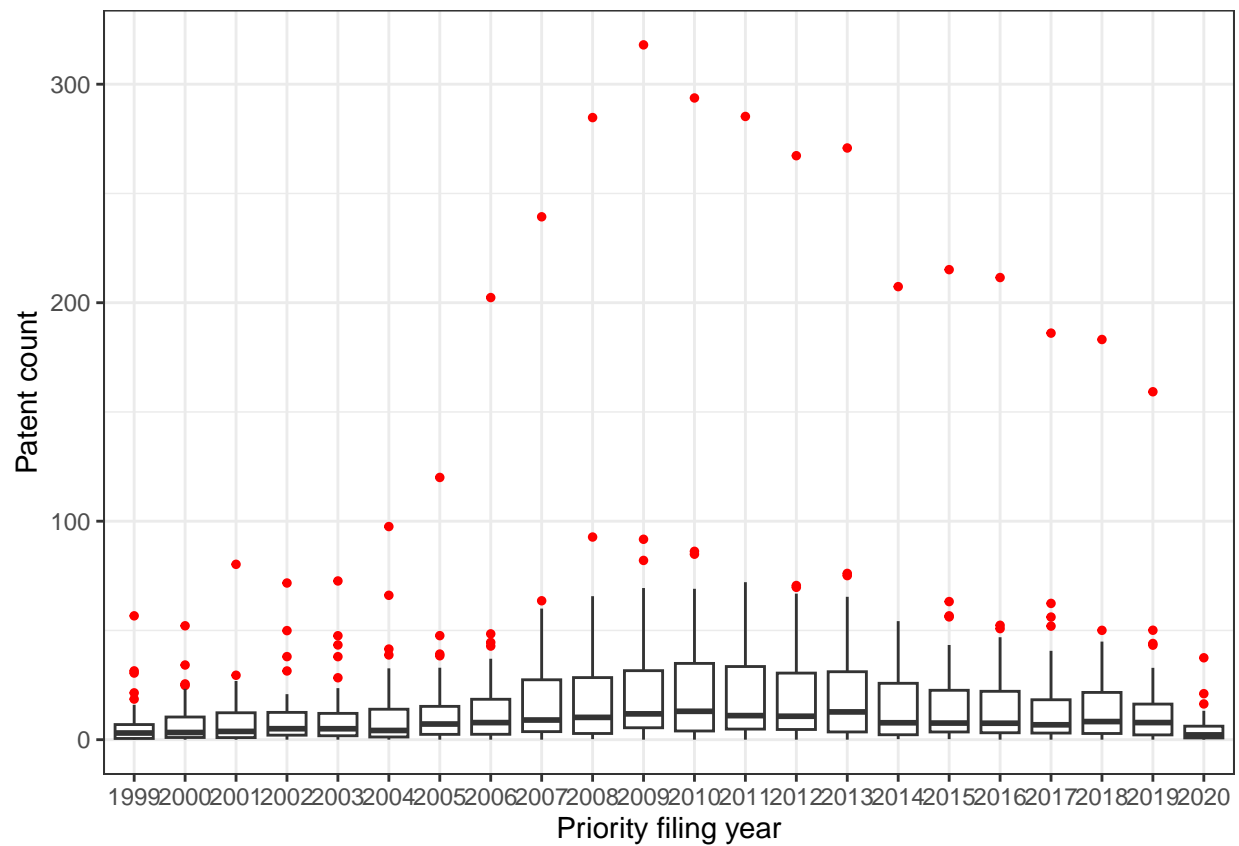
Here I assume that years that are not present = years with no patents filed under PCT

### Filter for specific time frame

## Summary statistics

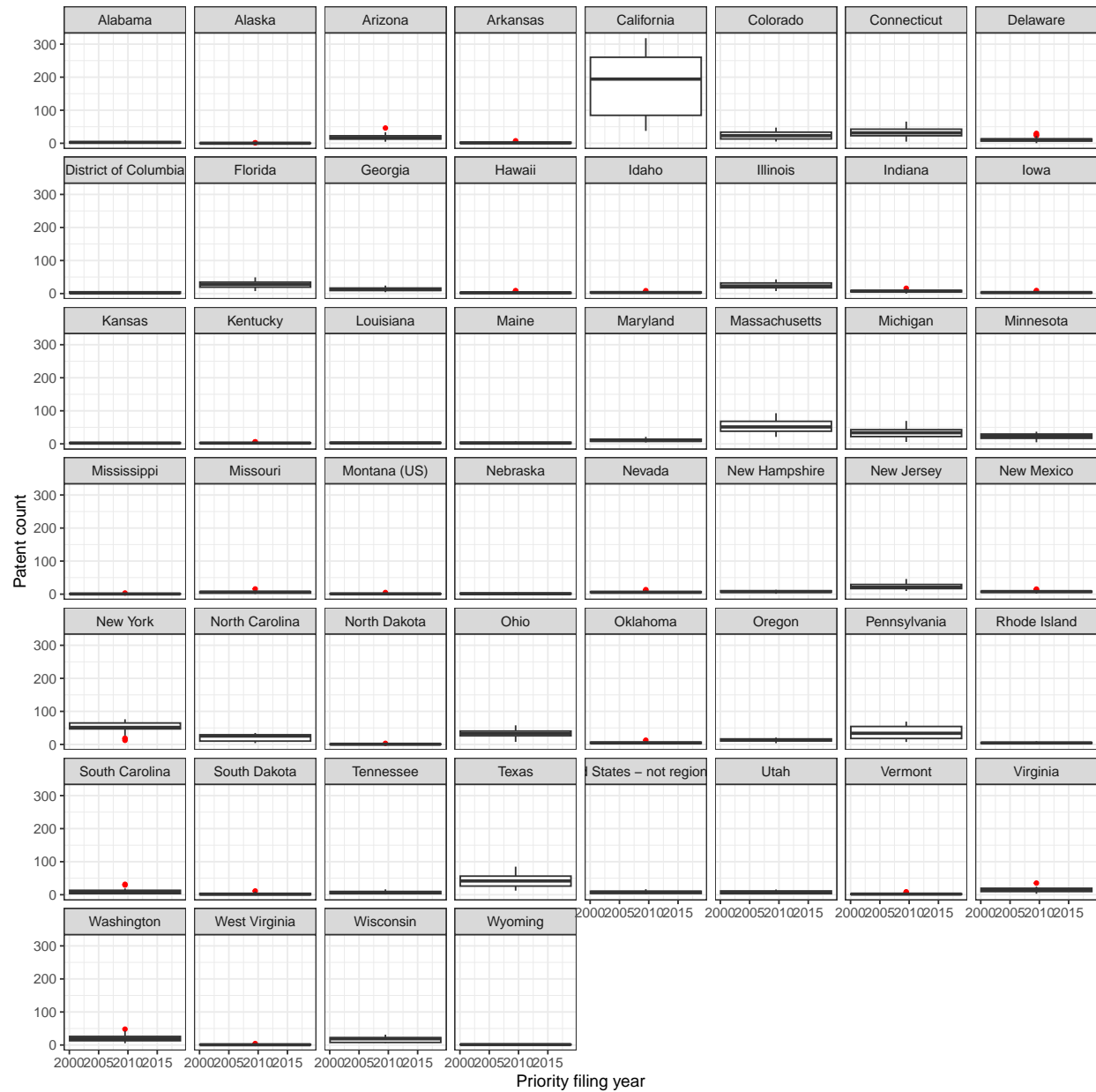
```
## # A tibble: 14 x 2
##   name                value
##   <chr>              <dbl>
## 1 prio_year_min      1999
## 2 prio_year_median  2010.
## 3 prio_year_mean    2010.
## 4 prio_year_stdev    6.35
## 5 prio_year_q25      2004
## 6 prio_year_q75      2015
## 7 prio_year_max      2020
## 8 patent_count_min    0
## 9 patent_count_median 6.94
## 10 patent_count_mean  16.2
## 11 patent_count_stdev 30.2
## 12 patent_count_q25   2.26
## 13 patent_count_q75   19.3
## 14 patent_count_max   318.
```

### Boxplot - by year



### Boxplot - by US State

```
## Warning: Continuous x aesthetic
## i did you forget 'aes(group = ...)'?
```



## Add controls

### GDP - 1999-2020

Real GDP (millions of chained 2012 dollars)

Unit: Millions of chained 2012 dollars

Source: BEA (Bureau of Economic Analysis)

### Population - 1999-2020

Resident population including Armed Forces

Unit of Measure: thousand

Sources: Table by the U.S. Energy Information Administration, based on population data from the U.S. Census Bureau

## **R&D - 1999-2019**

### **R&D and R&D Plant Obligations by U.S. State**

The amount of federal obligations for R&D and R&D plant to U.S. performers by state. Obligations represent the amounts for orders placed, contracts awarded, services received, and similar transactions during a given period, regardless of when the funds were appropriated.

Deflator: Current Dollars

Unit of Measure: Thousands of Dollars

SOURCE: National Center for Science and Engineering Statistics, Survey of Federal Funds for Research and Development.

## **Crude oil prices 1999-2020**

### **Domestic Crude Oil First Purchase Prices by Area**

Unit: Dollars per Barre

Definitions:

- Domestic: Crude oil produced in the U.S. or from its “outer continental shelf” as defined in 43 U.S.C. 1331.
- Crude Oil: A mixture of hydrocarbons that exists in liquid phase in natural underground reservoirs and remains liquid at atmospheric pressure after passing through surface separating facilities. Liquids produced at natural gas processing plants are excluded. Crude oil is refined to produce a wide array of petroleum products, including heating oils; gasoline, diesel and jet fuels; lubricants; asphalt; ethane, propane, and butane; and many other products used for their energy or chemical content.
- First purchase price: An equity (not custody) transaction involving an arms-length transfer of ownership of crude oil associated with the physical removal of the crude oil from a property (lease) for the first time. A first purchase normally occurs at the time and place of ownership transfer where the crude oil volume sold is measured and recorded on a run ticket or other similar physical evidence of purchase. The reported cost is the actual amount paid by the purchaser, allowing for any adjustments (deductions or premiums) passed on to the producer or royalty owner.

Numbers are aggregated by Petroleum Administration for Defense District (PADD):

- PADD 1 (East Coast):
  - PADD 1A (New England): Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont.
  - PADD 1B (Central Atlantic): Delaware, District of Columbia, Maryland, New Jersey, New York, Pennsylvania.
  - PADD 1C (Lower Atlantic): Florida, Georgia, North Carolina, South Carolina, Virginia, West Virginia.
- PADD 2 (Midwest): Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, Oklahoma, South Dakota, Tennessee, Wisconsin.
- PADD 3 (Gulf Coast): Alabama, Arkansas, Louisiana, Mississippi, New Mexico, Texas, Federal Off-shore Gulf.

- PADD 4 (Rocky Mountain): Colorado, Idaho, Montana, Utah, Wyoming.
- PADD 5 (West Coast): Alaska (North Slope and Other Mainland), Arizona, California, Hawaii, Nevada, Oregon, Washington, Federal Offshore California.

Source: U.S. Energy Information Administration, ([link to page](#))

#### **Alternatives:**

- From Energy Information Administration, there is a csv with a lot of information, by State and by year, on all kinds of energy consumption/production, including:
  - total energy average price in Dollars per million Btu
  - total energy consumption in Billion Btu
  - Renewable energy total consumption in Billion Btu
  - Crude oil production (including lease condensate) in Billion Btu
  - Crude oil consumed by the industrial sector in Billion Btu
  - Others...

#### **HDD & CDD**