Tx_Housing_ggplot2

Mathias

17/11/2020

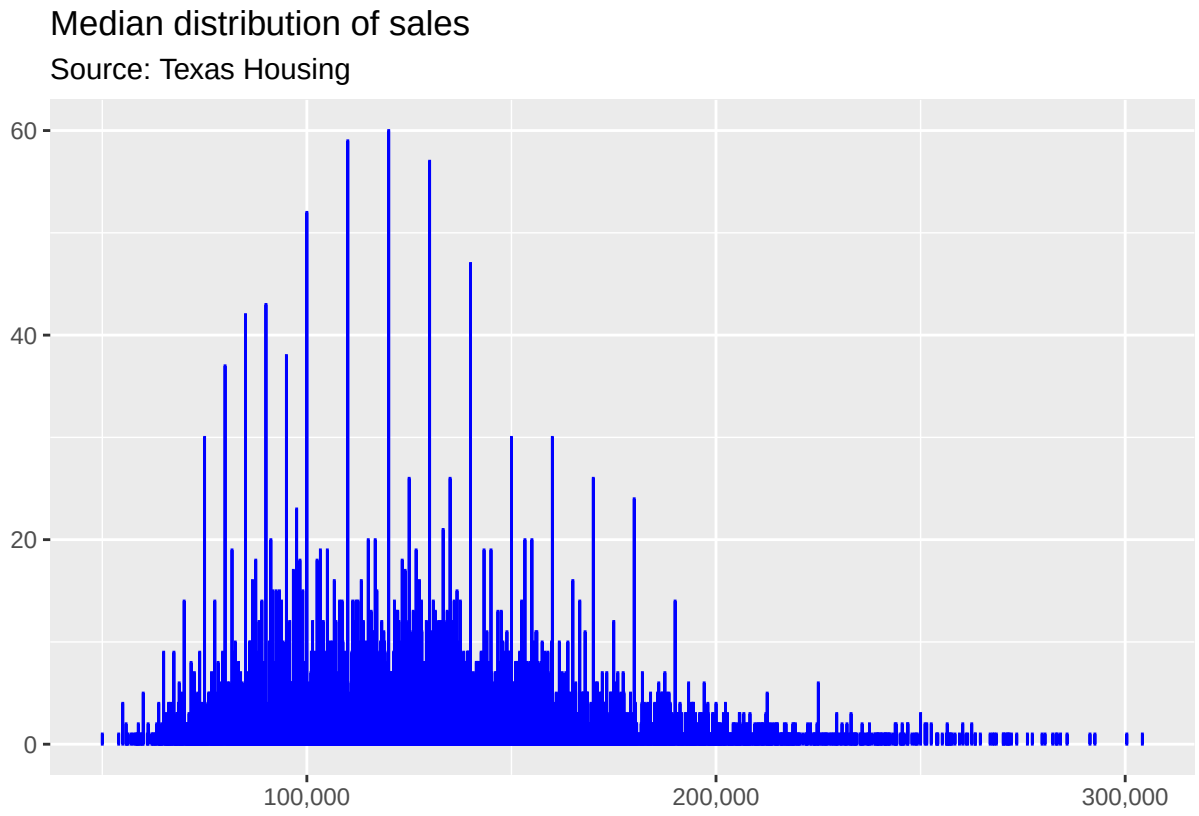
```
library(ggplot2)
library(scales)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
dat = read.csv('txhousing.csv')
dat$yearmonth = paste(dat$year, dat$month, sep = "/")
dat <- dat %>%
  mutate(city = as.factor(city),
         year = as.factor(year),
         month = as.factor(month),
         yearmonth = as.factor(yearmonth),
         sales = as.numeric(sales),
         volume = as.numeric(volume),
         mmedian = as.numeric(median),
         listings = as.numeric(listings))
```

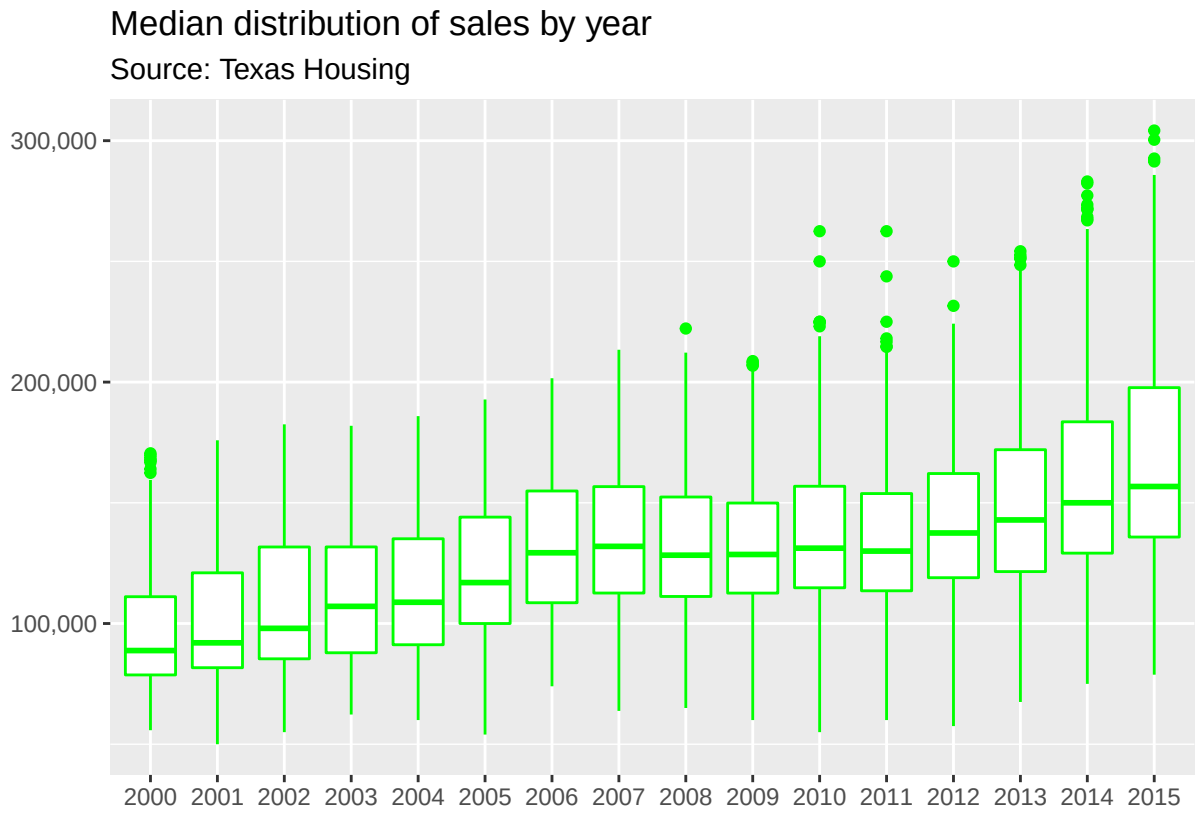
1 Distribution de la variable median

```
ggplot(data=dat) +
  geom_bar(mapping=aes(x=median), color="blue", width = 10) +
  ggtitle("Median distribution of sales", subtitle="Source: Texas Housing") +
  ylab(label = "") +
  xlab(label="") +
  scale_x_continuous(labels = comma)
```



2 Idem, en fonction de l'année

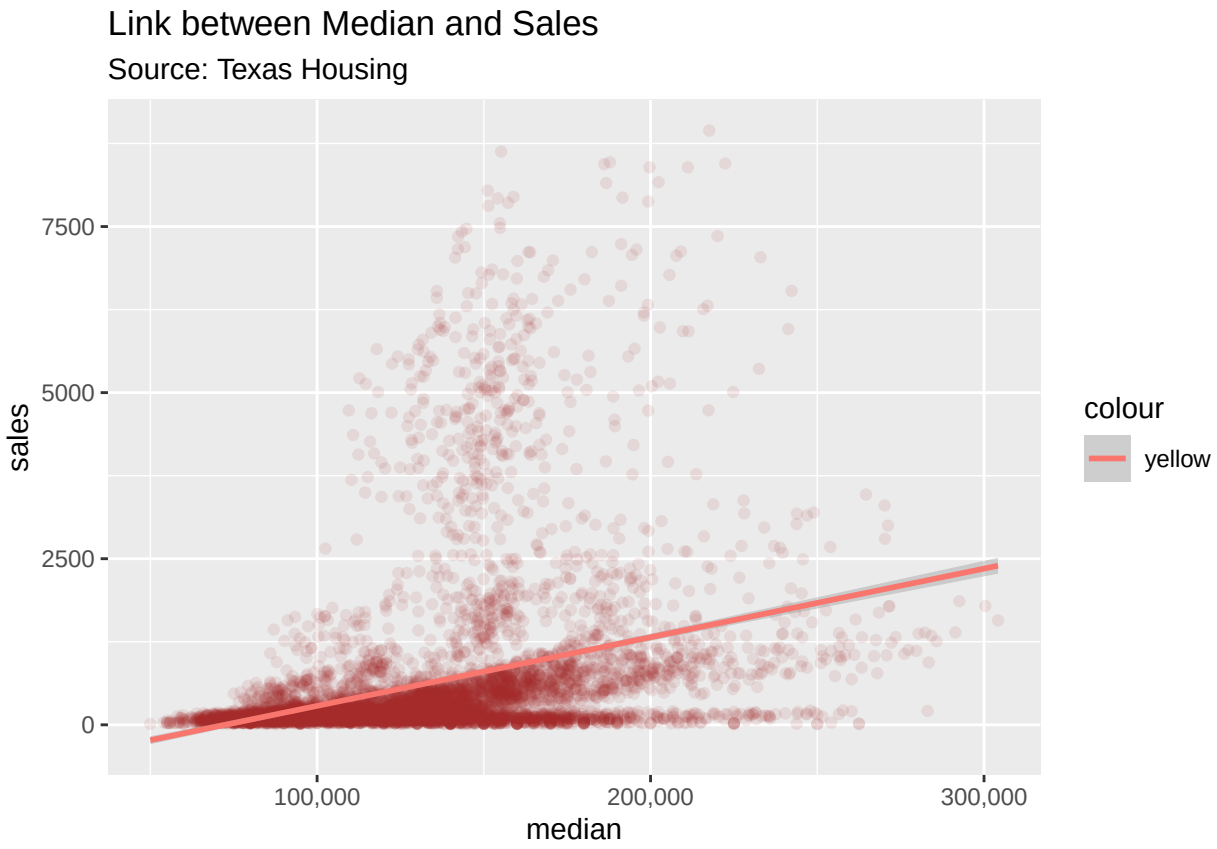
```
ggplot(data=dat) +
  geom_boxplot(mapping=aes(x=year, y = median), color="green") +
  ggtitle("Median distribution of sales by year", subtitle="Source: Texas Housing") +
  ylab(label = "") +
  xlab(label="") +
  scale_y_continuous(labels = comma)
```



3 Lien entre median et sales

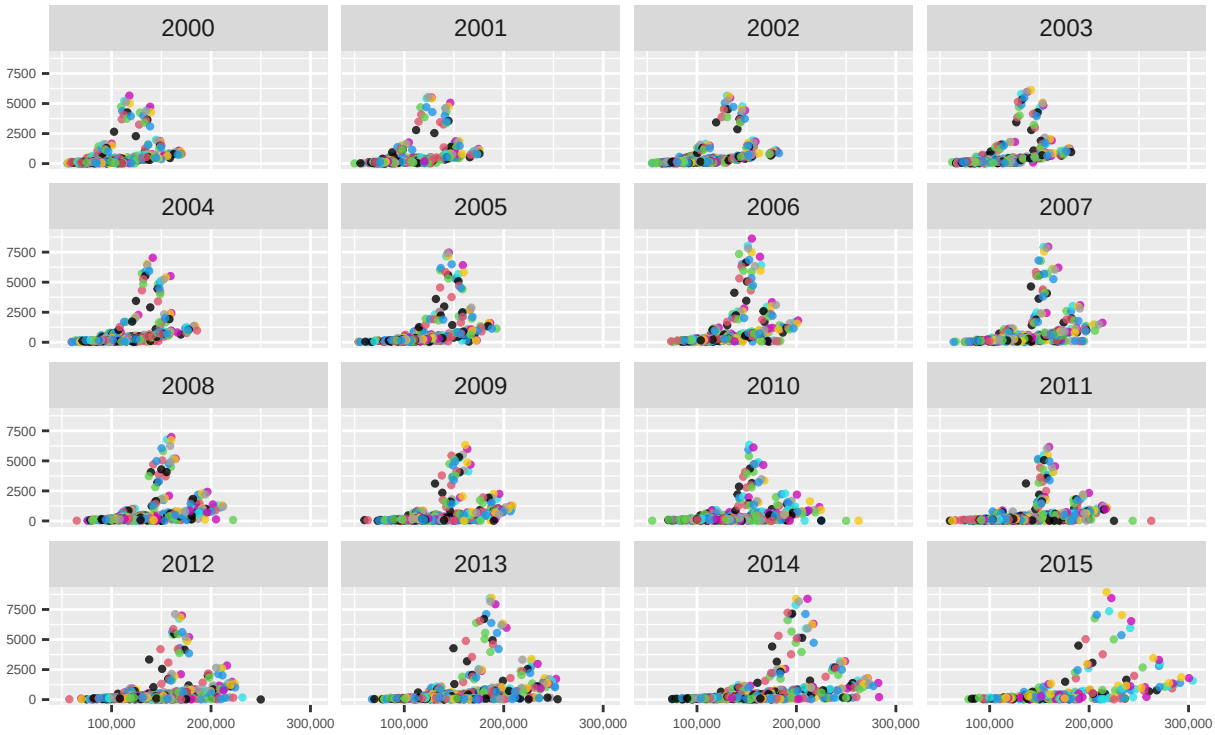
```
ggplot(data=dat) +
  geom_point(mapping=aes(x=median, y=sales), color="brown", alpha=0.1) +
  ggtitle("Link between Median and Sales", subtitle="Source: Texas Housing") +
  scale_x_continuous(labels = comma) +
  geom_smooth(aes(y=sales, x=median, color="yellow"), method = lm ) +
  theme()
```

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



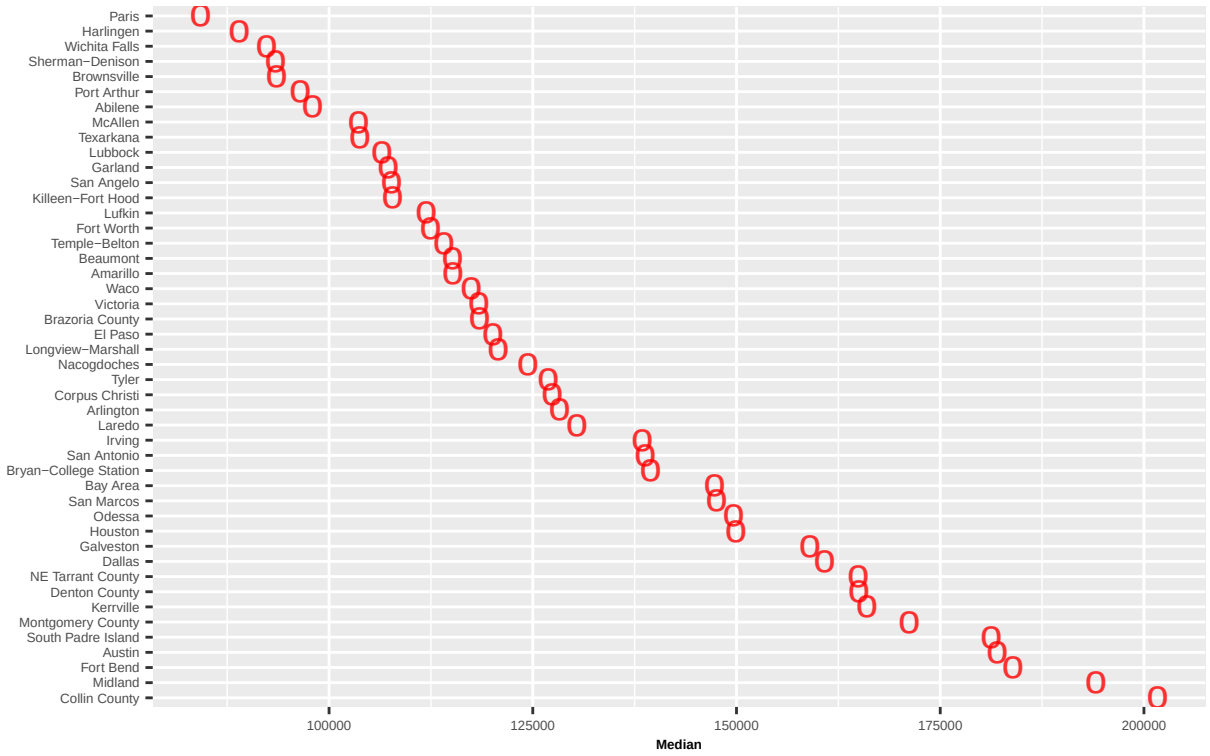
4 Idem, en fonction du mois et de l'année

```
ggplot(data=dat) +
  geom_point(mapping=aes(y=sales, x = median), color =dat$month, size=0.8, alpha=0.8) +
  ggtitle("Median distribution of sales by year and month", subtitle="Source: Texas Housing") +
  ylab(label = "") +
  xlab(label="") +
  scale_x_continuous(labels = comma) +
  theme(axis.text=element_text(size=5),
        axis.title=element_text(size=5,face="bold")) +
  facet_wrap(~year)
```



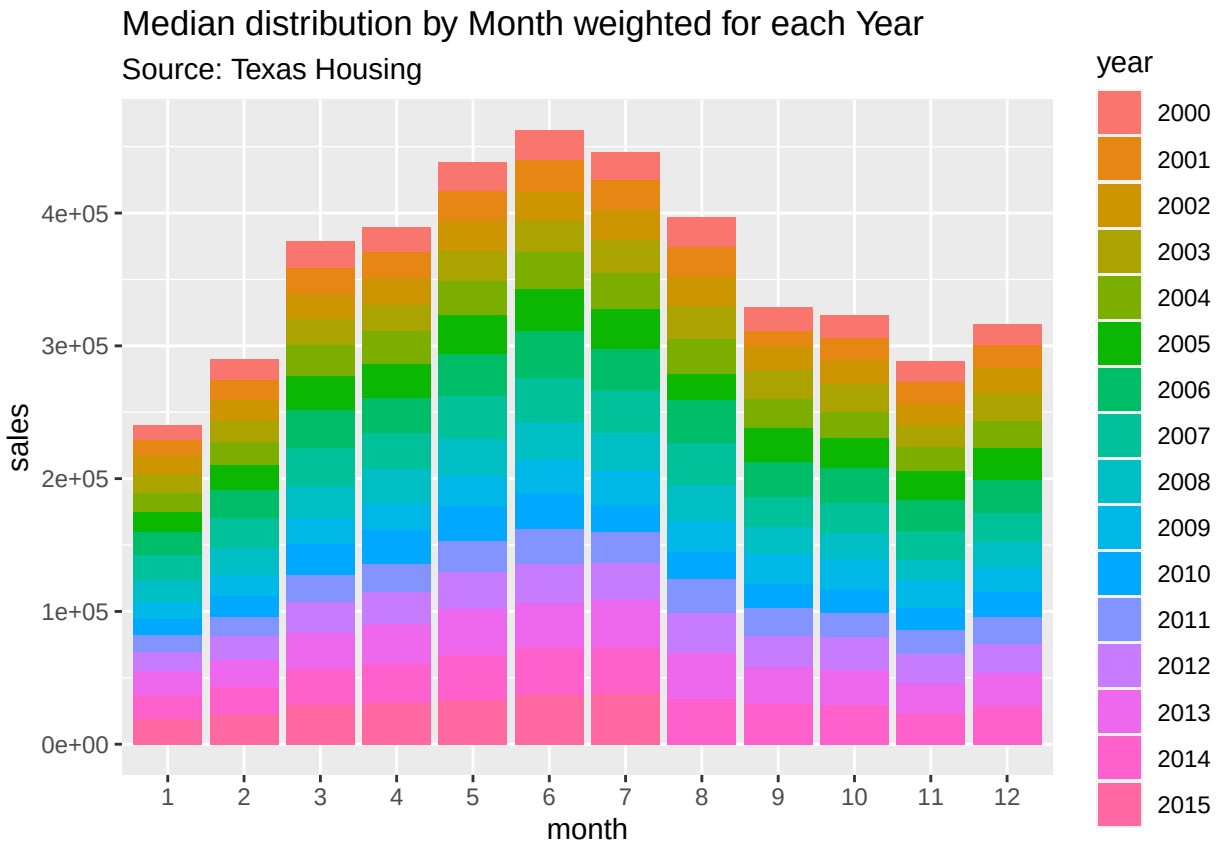
5 Evolution globale de median pour l'ensemble des villes

```
dat.aggcity <- aggregate(data = dat, . ~city , mean)
ggplot(data=dat.aggcity) +
  geom_point(mapping=aes(y= reorder(city, -median), x = median), color="red", size = 5, shape = "o", al
  ggtitle("Median distribution by cities", subtitle="Source: Texas Housing") +
  ylab(label = "") +
  xlab(label="Median" ) +
  theme(axis.text=element_text(size=5),
        axis.title=element_text(size=5,face="bold"))
```



6 Idem, par mois avec une couleur par année

```
ggplot(data=dat, na.rm = TRUE) +
  geom_bar(mapping=aes(x= month, y = sales, fill=year), stat="identity") +
  ggtitle("Median distribution by Month weighted for each Year", subtitle="Source: Texas Housing")
```


Median distribution by Month weighted for each Year
Source: Texas Housing
sales
4e+05
3e+05
2e+05
1e+05
0e+00
1
2
3
4
5
6
7
8
9
10
11
12
month
year
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015