

main2

2024-01-21

Seconda parte del progetto

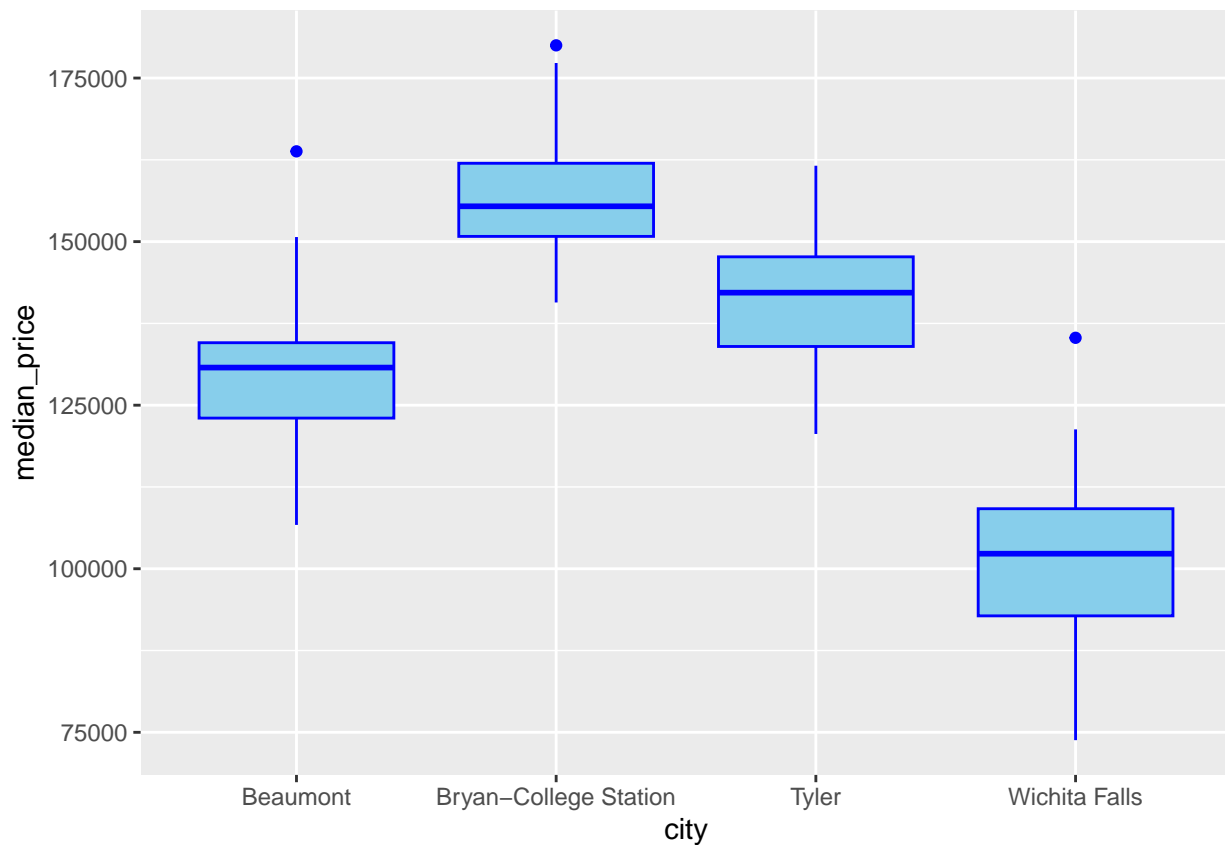
```
#install.packages("ggplot2")
#install.packages("dplyr")

library("dplyr")
library("ggplot2")
```

- 1) Utilizza i boxplot per confrontare la distribuzione del prezzo mediano delle case tra le varie città.
Commenta il risultato.

```
RealEstateTexax_Dataframe = read.csv("Real Estate Texas.csv")
attach(RealEstateTexax_Dataframe)

ggplot(data = RealEstateTexax_Dataframe) +
  geom_boxplot( aes(x = city, y = median_price ), fill="skyblue", color = "blue", lwd = 0.5)
```

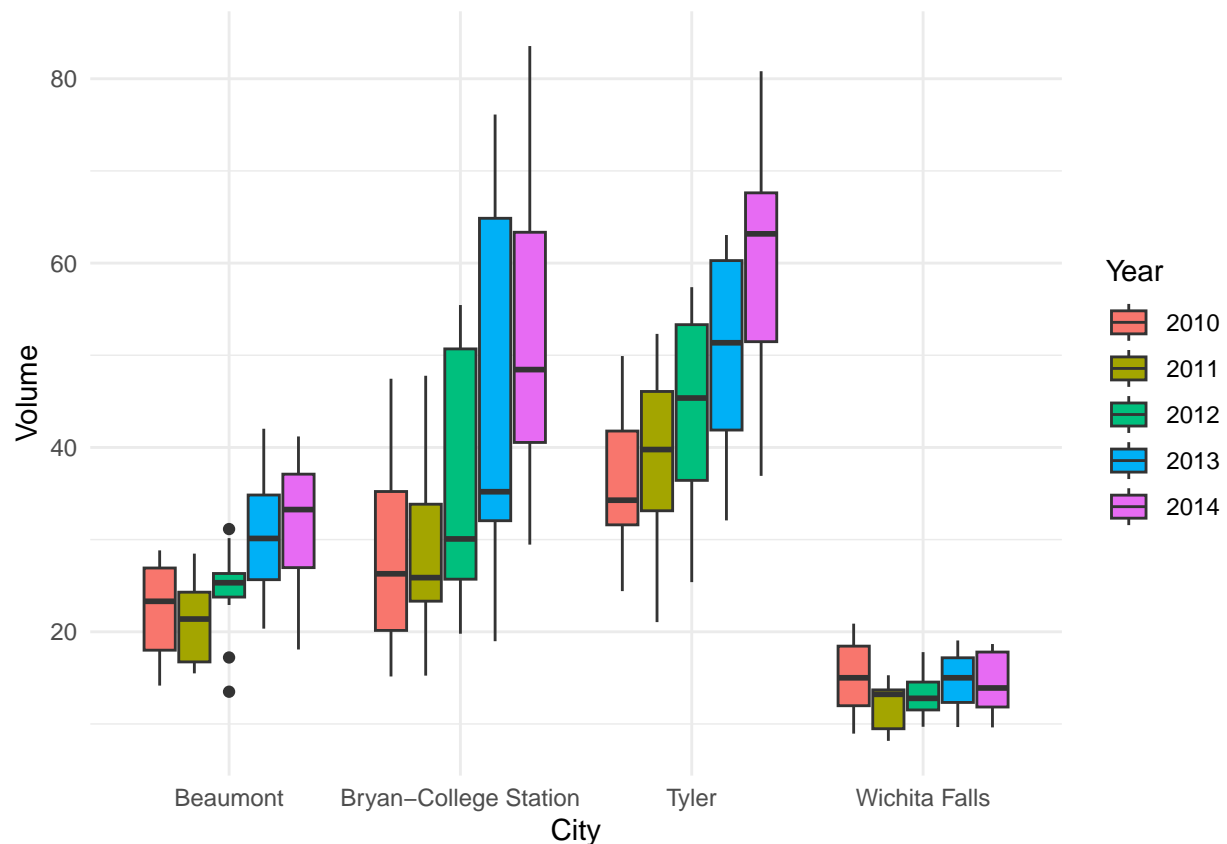


1.1) Le case con il median price più alto si trovano a Bryan-College Station e quelle con quello più basso a Wichita Falls.

1.2) Le case con i median price più variabili sembrano trovarsi a Wichita Falls, dato ricavato osservando la maggior ampiezza del range interquantile rispetto agli altri e anche dalla maggior lunghezza delle code, considerando anche gli outliers. Al contrario, quelle con i median price meno variabili a Bryan-College Station.

2) Utilizza i boxplot o qualche variante per confrontare la distribuzione del valore totale delle vendite tra le varie città ma anche tra i vari anni. Qualche considerazione da fare?

```
ggplot(data = RealEstateTexax_Dataframe) +  
  geom_boxplot(aes(x = city, y = volume, fill = as.factor(year))) + #as.factor() -> per trasforma  
  labs(x = "City", y = "Volume", fill = "Year") +  
  theme_minimal()
```



2.1) il volume totale delle vendite a Wichita Falls è il più basso tra tutti e non presenta trend, si è infatti mosso all'interno di un range orizzontale. Anche la varianza è rimasta costante negli anni.

2.2) per le altre città invece si può notare un movimento direzionale, meno evidente per Beaumont, mentre molto in vista per le restanti 2.

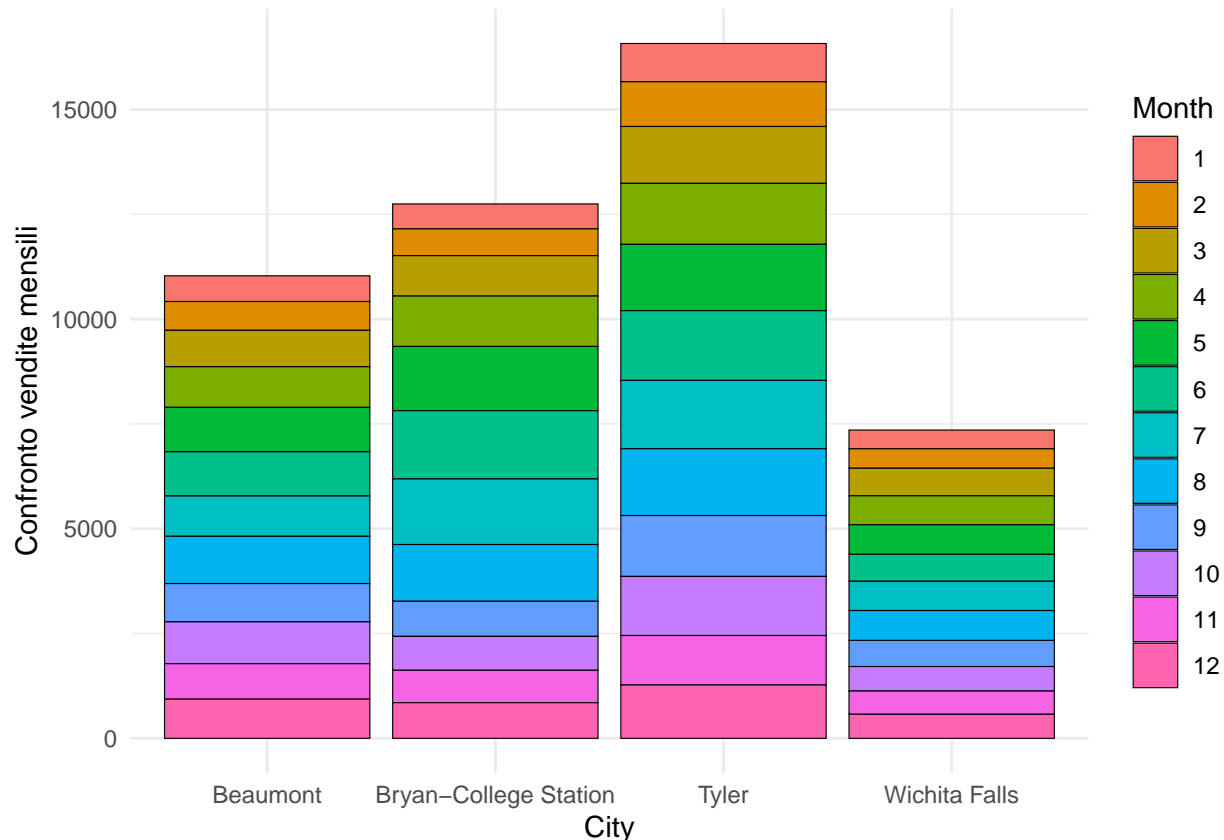
2.3) Oltre ad avere un trend positivo per quanto riguarda il volume delle vendite, Bryan-College Station e Tyler presentano anche un trend positivo per quanto riguarda la variabilità, la quale infatti è aumentata al passare degli anni.

2.4) Dal grafico si può giungere quindi alla conclusione che negli ultimi anni l'interesse per l'acquisto di una casa a Bryan-College Station o a Tyler è aumentato in maniera significativa, leggermente a Beaumont e per nulla a Wichita falls.

- 3) Usa un grafico a barre sovrapposte per confrontare il totale delle vendite nei vari mesi, sempre considerando le città. Prova a commentare ciò che viene fuori. Già che ci sei prova anche il grafico a barre normalizzato. Consiglio: Stai attento alla differenza tra `geom_bar()` e `geom_col()`. PRO LEVEL: cerca un modo intelligente per inserire ANCHE la variabile Year allo stesso blocco di codice, senza però creare accrocchi nel grafico.

```
totale_vendite_mensili <- RealEstateTexax_Dataframe %>%
  group_by(city, month) %>%
  summarise(totale_vendite = sum(sales, .groups=month, na.rm = TRUE))

ggplot(data = totale_vendite_mensili) +
  geom_bar(aes(x = city, y = totale_vendite_mensili$totale_vendite, fill = as.factor(month)),
    position = "stack", stat = "identity", col = "black", lwd = 0.2) +
  labs(x = "City", y = "Confronto vendite mensili") +
  scale_fill_discrete(name = "Month") +
  theme_minimal()
```



Dal grafico a barre sovrapposte si può notare come le vendite nella città di Beaumont e Wichita Falls siano state più o meno costanti durante tutti i mesi, mentre per la città di Bryan-College Station e (in maniera minore Tyler) sembra che le vendite siano maggiori durante i mesi primaverili ed estivi piuttosto che negli altri.

Inoltre il numero totale di vendite è maggiore nella città di Tyler e a seguire in Bryan-College Station, Beaumont e Wichita Falls.

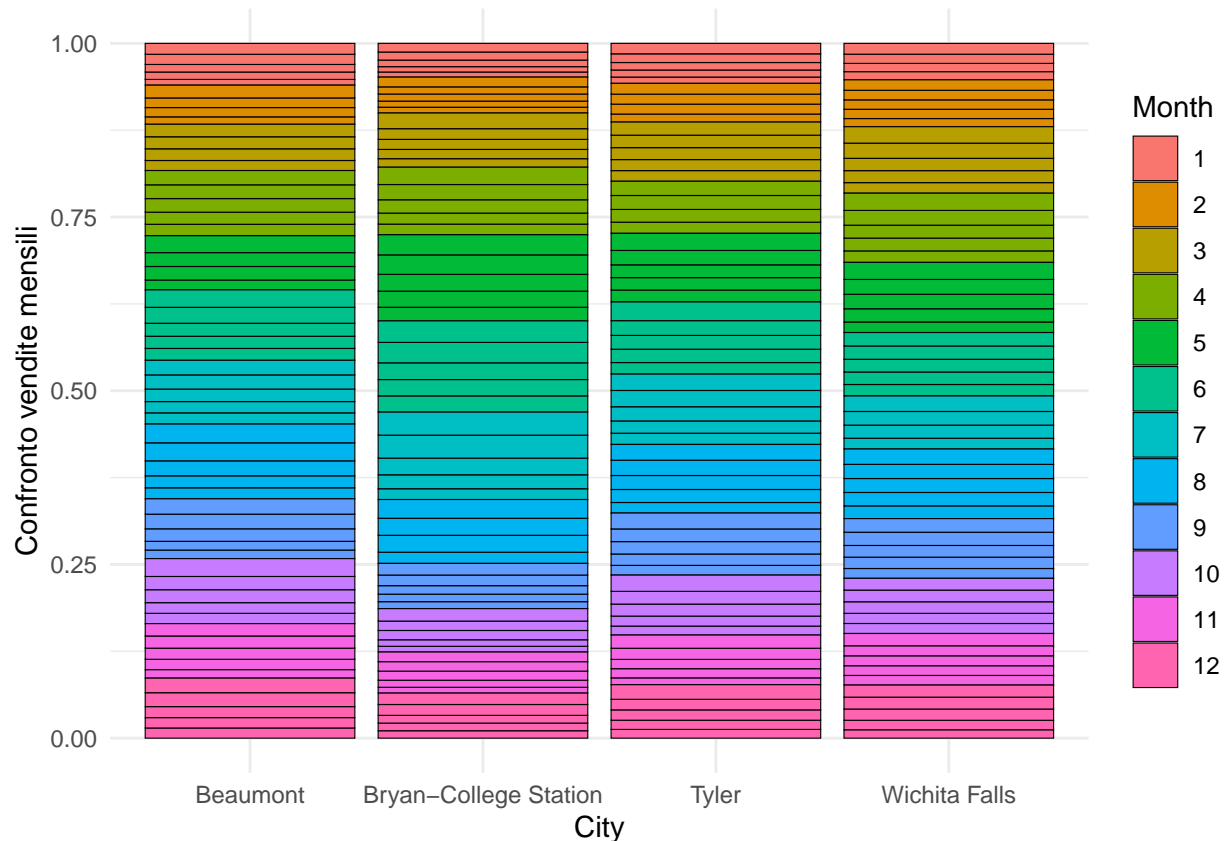
grafico a barre sovrapposte normalizzato:

```
ggplot(data = totale_vendite_mensili) +
  geom_bar( aes(x = city, y = totale_vendite_mensili$totale_vendite, fill = as.factor(month)),
    position = "fill", stat = "identity", col = "black", lwd = 0.2) +
  labs(x = "City", y = "Confronto vendite mensili") +
  scale_fill_discrete(name = "Month") +
  theme_minimal()
```



#PRO LEVEL

```
ggplot(data = RealEstateTexax_Dataframe["sales"]) +
  geom_bar( aes(x = city, y = sales, fill = as.factor(month), col = "grey", lwd = 1),
    position = "fill", stat = "sum", col = "black", lwd = 0.2) +
  labs(x = "City", y = "Confronto vendite mensili") +
  scale_fill_discrete(name = "Month") +
  theme_minimal()
```



- 4) Prova a creare un line chart di una variabile a tua scelta per fare confronti commentati fra città e periodi storici.

```
dati_Beaumont <- filter(RealEstateTexax_Dataframe, city == "Beaumont")
dati_Bryan_college <- filter(RealEstateTexax_Dataframe, city == "Bryan-College Station")
dati_Tyler <- filter(RealEstateTexax_Dataframe, city == "Tyler")
dati_Wichita <- filter(RealEstateTexax_Dataframe, city == "Wichita Falls")

dati_Beaumont <- dati_Beaumont %>%
  rename_with(~paste0("Beaumont_", .), everything())

dati_Bryan_college <- dati_Bryan_college %>%
  rename_with(~paste0("Bryan_college_", .), everything())

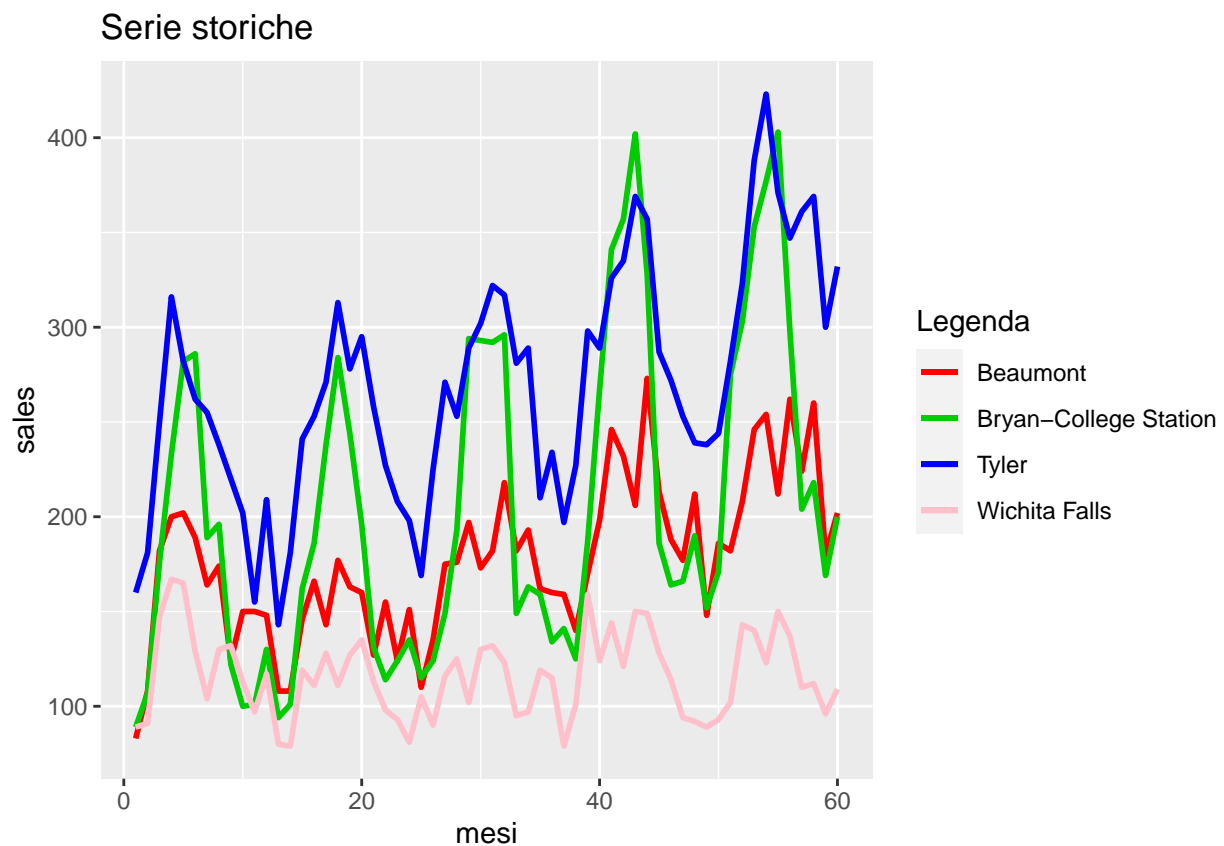
dati_Tyler <- dati_Tyler %>%
  rename_with(~paste0("Tyler_", .), everything())

dati_Wichita <- dati_Wichita %>%
  rename_with(~paste0("Wichita_", .), everything())

new_dataframe = as.data.frame(cbind(dati_Beaumont, dati_Bryan_college, dati_Tyler, dati_Wichita))

attach(new_dataframe)
```

```
ggplot(data = new_dataframe)+
  geom_line(aes(x = seq(1,60,1), y = Beaumont_sales, col="Beaumont"), lwd=1)+
  geom_line(aes(x = seq(1,60,1), y = Bryan_college_sales, col="Bryan-College Station"), lwd=1)+
  geom_line(aes(x = seq(1,60,1), y = Tyler_sales, col="Tyler"), lwd=1)+
  geom_line(aes(x = seq(1,60,1), y = Wichita_sales, col="Wichita Falls"), lwd=1)+
  labs(x="mesi",
       y="sales",
       title="Serie storiche")+
  scale_color_manual(
    name = 'Legenda',
    breaks = c('Beaumont', 'Bryan-College Station', 'Tyler', 'Wichita Falls'),
    values = c("red", "green3", "blue", "pink"),
    labels = c('Beaumont', 'Bryan-College Station', 'Tyler', 'Wichita Falls'))
```



Wichita Falls non ha subito incrementi nelle vendite negli anni e la sua bivaribilità è rimasta la stessa. Beaumont ha visto un aumento nelle vendite col passare degli anni ma non della variabilità.

Al contratio Bryan-College Station, Tyler hanno visto, col passare degli anni, sia un aumento nella quantità delle vendite sia della loro stessa variabilità.

città con meno vendite (in media): Wichita Falls città con più vendite (in media): Tyler