

Analisi esplorativa del mercato immobiliare del Texas

Il **dataset** contiene dei dati riguardanti le vendite di immobili in Texas

city: città

year: anno di riferimento

month: mese di riferimento (codificato con numeri da 1 a 12)

sales: numero totale di vendite

volume: valore totale delle vendite in milioni di dollari

median_price: prezzo mediano di vendita in dollari

listings: numero totale di annunci attivi

months_inventory: quantità di tempo necessaria per vendere tutte le inserzioni correnti al ritmo attuale delle vendite, espresso in mesi.

Tipo di variabili

VARIABILI QUANTITATIVE

Continua, su scala di intervalli: **year**

Discrete, su scala di rapporti:
sales, listings

Continue, su scala di rapporti:
volume, median_price,
months_inventory

VARIABILI QUALITATIVE

Su scala nominale: **city**

Ciclica: **month**

Indici di posizione: moda

year	character [5]	'2010' '2011' '2012' '2013' '2014'
city	character [4]	'Beaumont' 'Bryan-College Station' 'Tyler' 'Wichita Falls'
months_inventory	character [1]	' 8.1'
month	character [12]	' 1' ' 2' ' 3' ' 4' ' 5' ' 6' ...

Le variabili *city*, *year* e *month* hanno come moda tutti i loro possibili valori, che si presentano nel dataset con frequenze identiche

Indici di posizione: quartili e media

Minimo e massimo

year	Min. :2010
year	Max. :2014

listings	Min. : 743
listings	Median :1618
listings	Mean :1738
listings	Max. :3296
listings	3rd Qu.:2056
listings	1st Qu.:1026

median_price	Min. : 73800
median_price	Median :134500
median_price	Mean :132665
median_price	Max. :180000
median_price	3rd Qu.:150050
median_price	1st Qu.:117300

volume	Min. : 8.166
volume	Median :27.062
volume	Mean :31.005
volume	Max. :83.547
volume	3rd Qu.:40.893
volume	1st Qu.:17.660

sales	Min. : 79.0
sales	Median :175.5
sales	Mean :192.3
sales	Max. :423.0
sales	3rd Qu.:247.0
sales	1st Qu.:127.0

months_inventory	Min. : 3.400
months_inventory	Median : 8.950
months_inventory	Mean : 9.193
months_inventory	Max. :14.900
months_inventory	3rd Qu.:10.950
months_inventory	1st Qu.: 7.800

Indici di variabilità: range

year	double [1]	4
sales	double [1]	344
volume	double [1]	75.381
median_price	double [1]	106200
listings	double [1]	2553
months_inventory	double [1]	11.5

Indici di variabilità: range interquartile

sales	double [1]	120
volume	double [1]	23.2335
median_price	double [1]	32750
listings	double [1]	1029.5
months_inventory	double [1]	3.15

Indici di variabilità: varianza

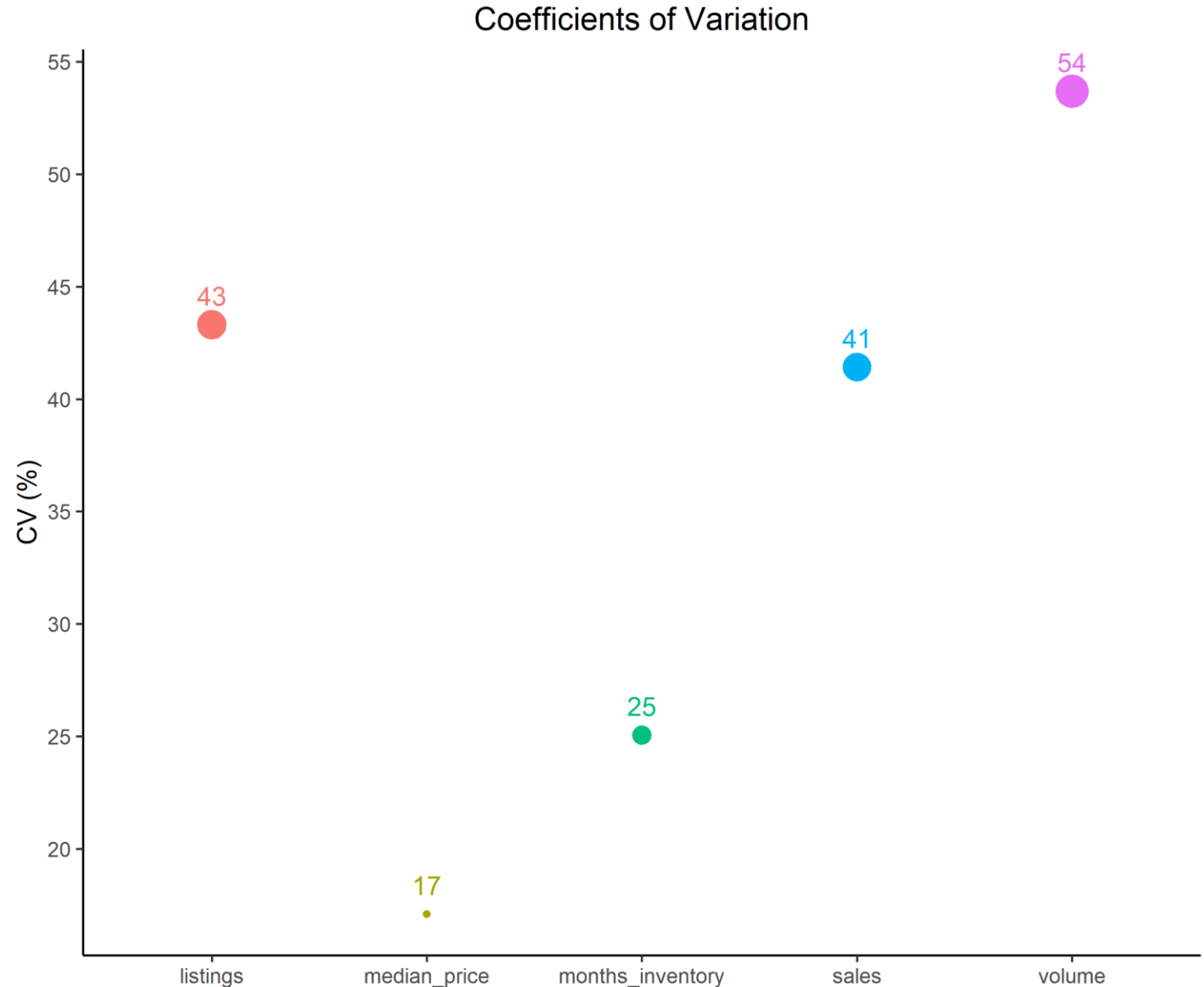
sales	double [1]	6344.3
volume	double [1]	277.2707
median_price	double [1]	513572983
listings	double [1]	566569
months_inventory	double [1]	5.306889

Indici di variabilità: deviazione standard

sales	double [1]	79.65111
volume	double [1]	16.65145
median_price	double [1]	22662.15
listings	double [1]	752.7078
months_inventory	double [1]	2.303669

Indici di variabilità: Coefficiente di Variazione

- La variabile con più alta variabilità è *volume*
- La variabile con minore variabilità è *median_price*



Indici di variabilità: indice di Gini

city	month
1	1

L'indice di Gini è pari a 1 per entrambe le variabili qualitative, *city* e *month*, per le quali ogni modalità ha pari frequenza

Indici di forma: asimmetria

<code>sales</code>	<code>double [1]</code>	0.718104
<code>volume</code>	<code>double [1]</code>	0.884742
<code>median_price</code>	<code>double [1]</code>	-0.3645529
<code>listings</code>	<code>double [1]</code>	0.6494982
<code>months_inventory</code>	<code>double [1]</code>	0.04097527

Si osserva che:

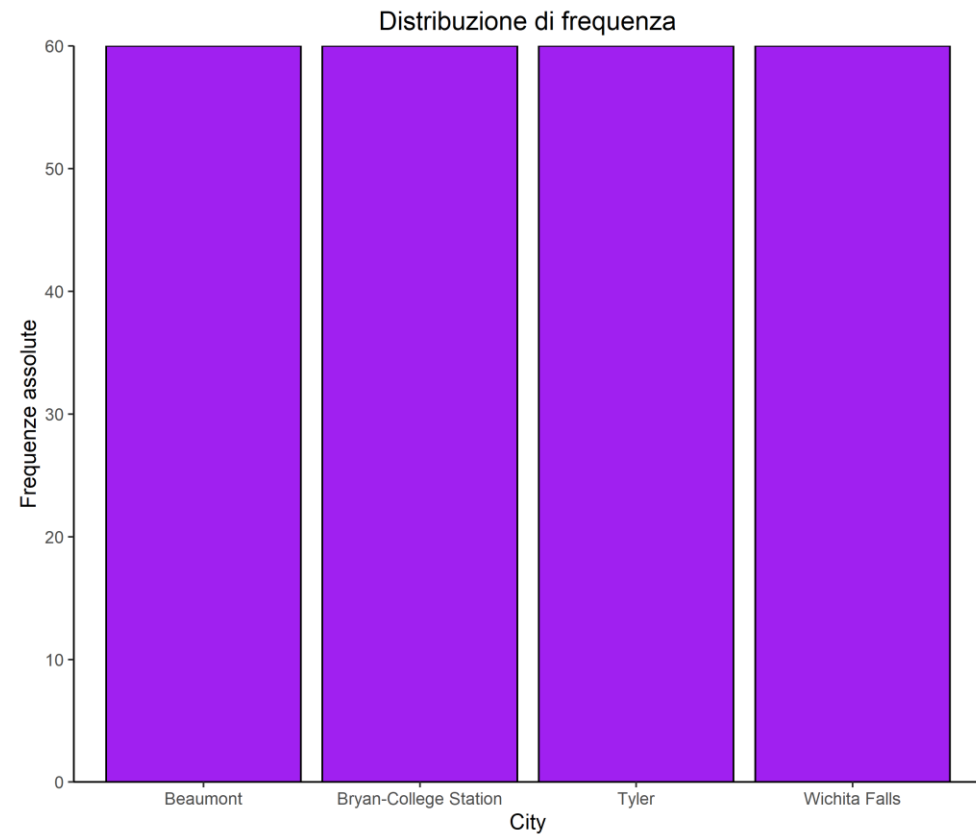
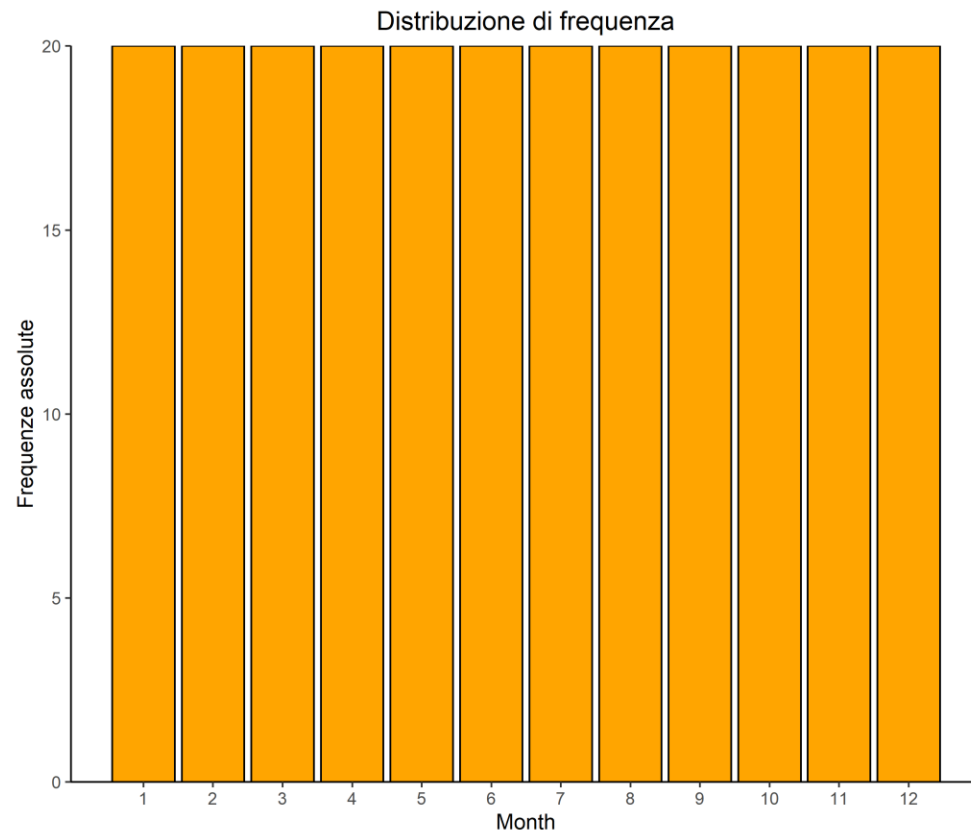
- la variabile *median_price* ha distribuzione asimmetrica negativa
- le variabili *sales*, *volume*, *listings* e *months_inventory* hanno distribuzione asimmetrica positiva
- la variabile più asimmetrica risulta *volume*, in quanto ha la skewness con modulo più elevato

Indici di forma: curtosi

sales	double [1]	2.686824
volume	double [1]	3.176987
median_price	double [1]	2.377038
listings	double [1]	2.20821
months_inventory	double [1]	2.825552

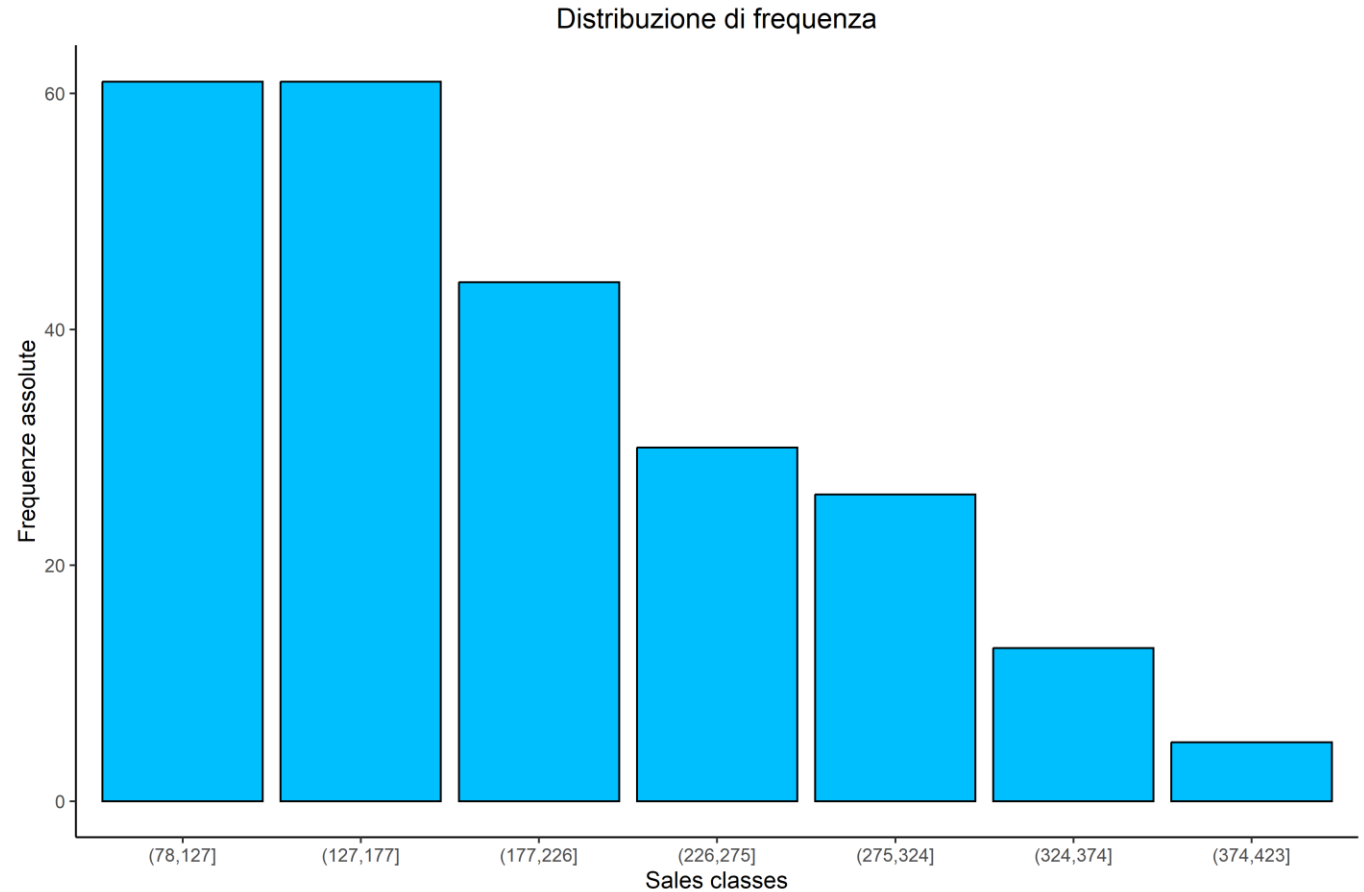
Tutte le variabili quantitative hanno distribuzione leptocurtica

Distribuzioni di frequenza



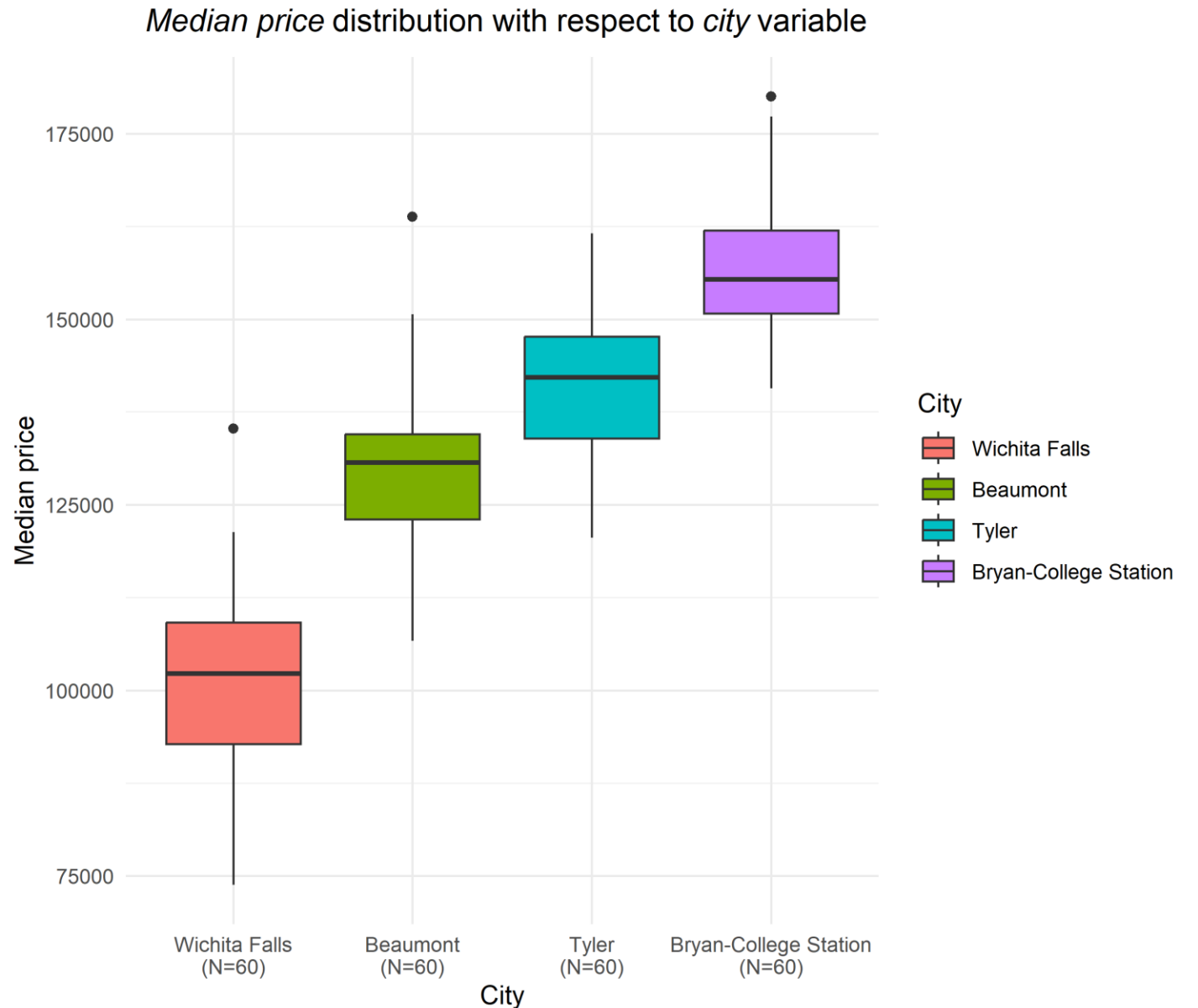
Distribuzione di frequenza della variabile *sales* suddivisa in classi

- La distribuzione di frequenza della variabile *sales_cl* risulta asimmetrica positiva, così come mostrato dal valore di skewness della variabile non suddivisa in classi.
- L'indice di Gini per la variabile suddivisa in classi è pari a 0.941, indicando che è presente più di una classe.



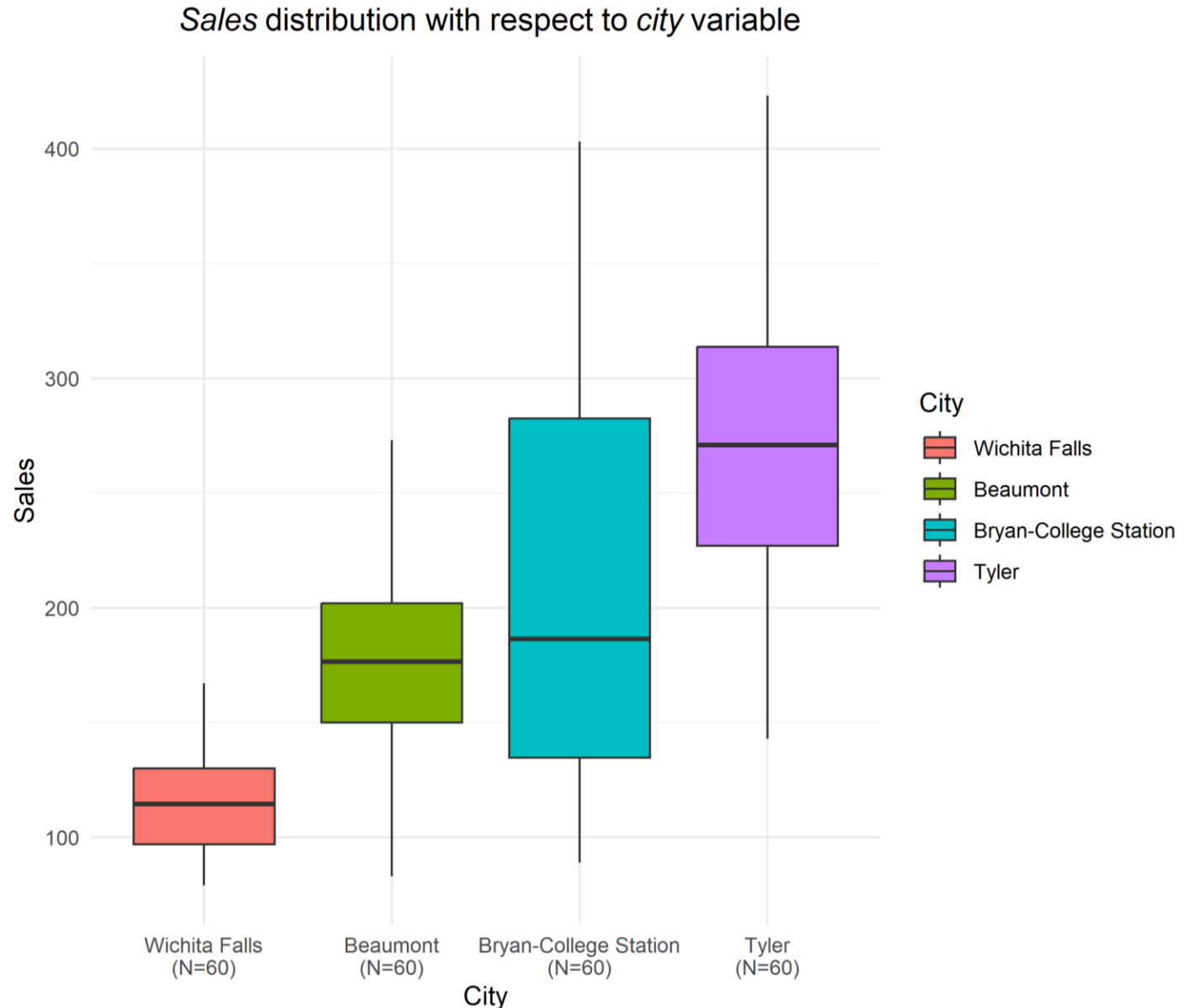
Boxplot della variabile *median_price*

- Il prezzo mediano di vendita ha una distribuzione con valori mediamente crescenti secondo l'ordine seguente: Wichita Falls, Beaumont, Tyler, Bryan-College Station.
- Prendendo a riferimento la città con valori più bassi, Wichita Falls, ed escludendo il suo outlier, le altre città hanno la maggior parte o la totalità dei valori più alti.
- Il prezzo mediano per ogni città ha un range che varia tra circa 40 k e 60 k.



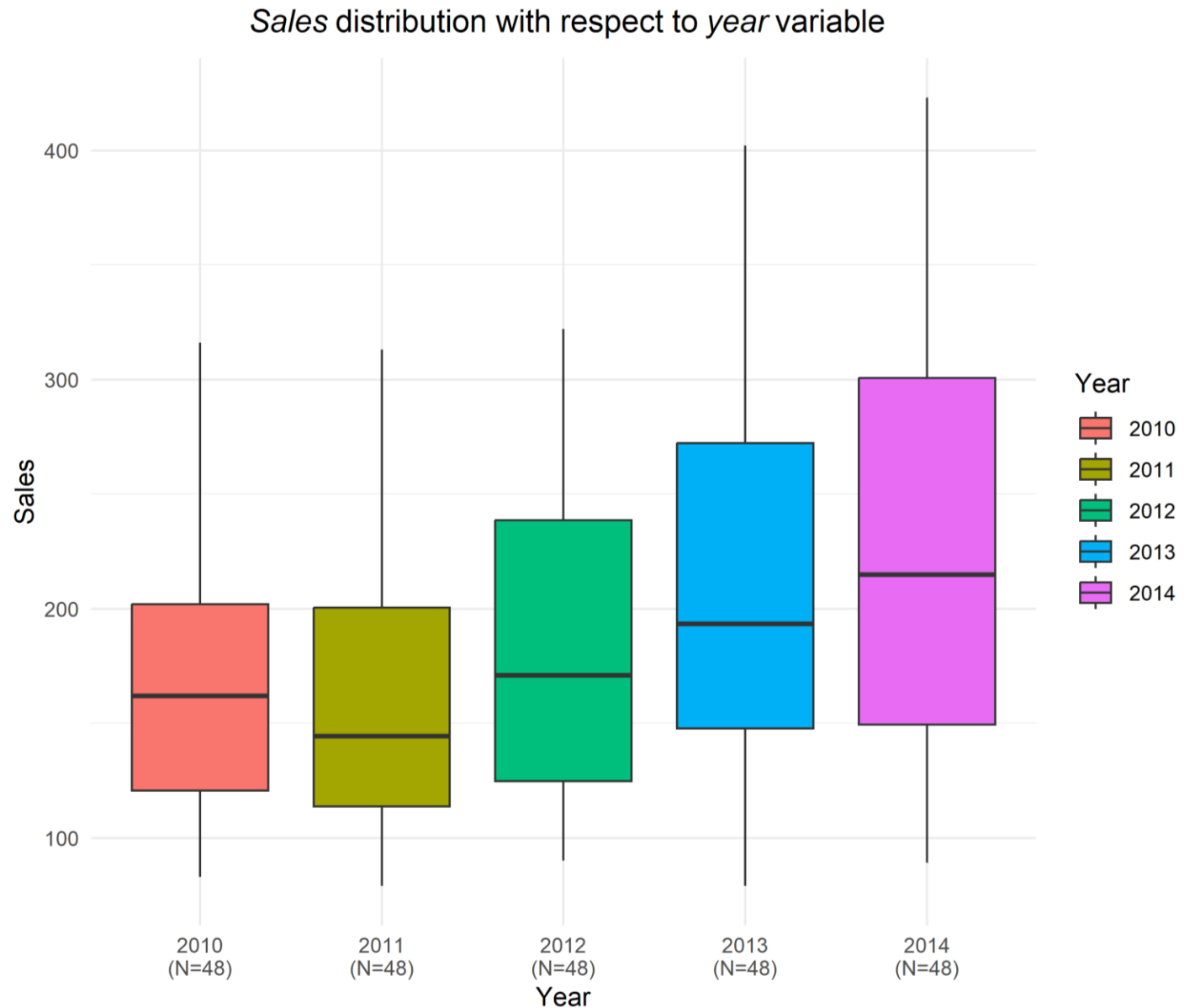
Boxplot della variabile *sales* rispetto alla variabile *city*

- Il numero di vendite mensili ha il range minore e valori più bassi a Wichita Falls. Il range diventa oltre il doppio e i valori più alti a Beaumont. Bryan-College Station e Tyler raggiungono vendite mensili ancora maggiori e in un range che è oltre il triplo rispetto a Wichita Falls, nonostante siano anche le città con prezzo mediano più alto.
- Il numero minimo di vendite è simile in tutte le città tranne Tyler, che è anche la città dove si raggiunge il numero mensile di vendite più alto.
- In generale quindi le vendite mensili vanno meglio a Tyler, seguita da Bryan-College Station, Beaumont e infine Wichita Falls.



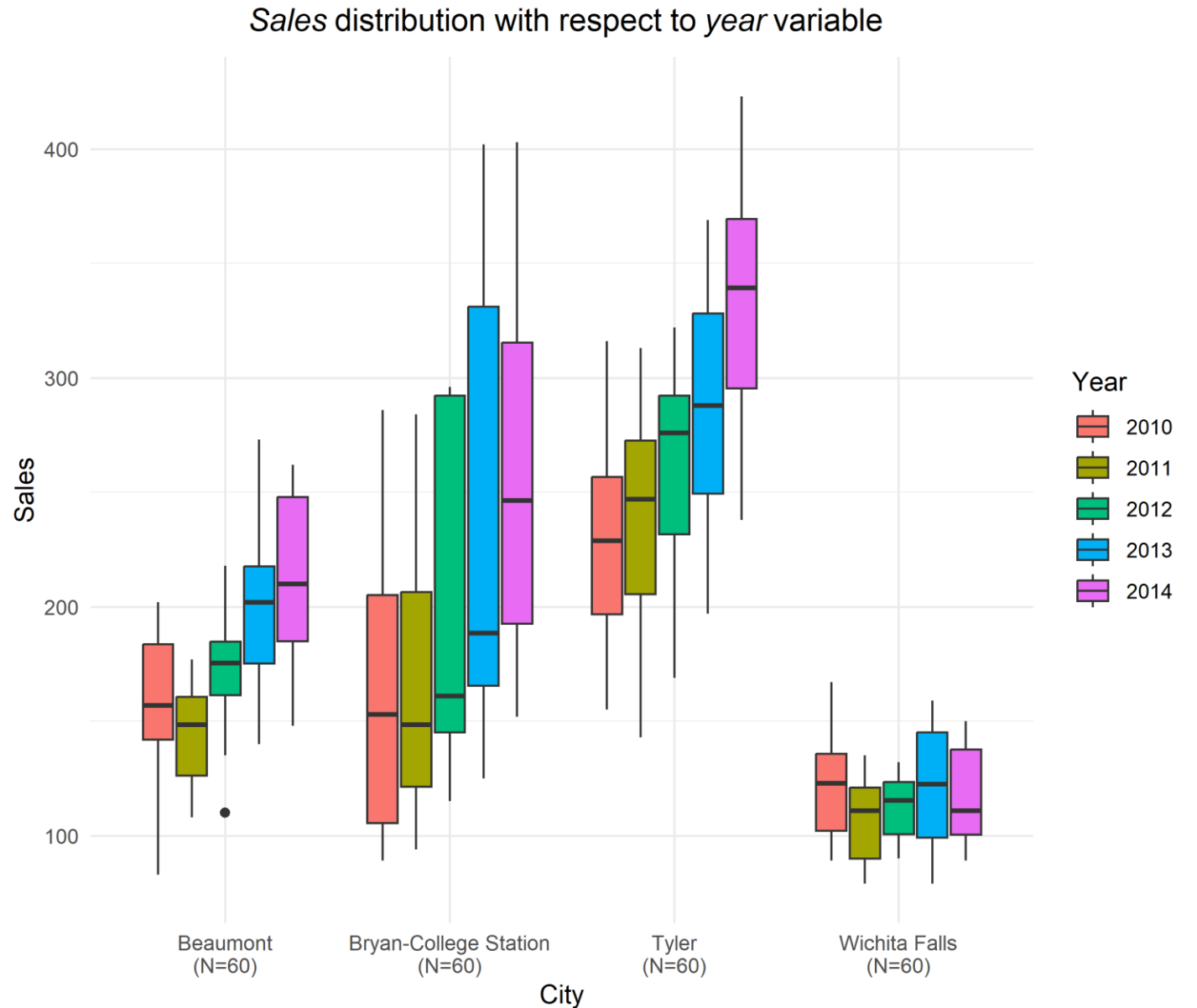
Boxplot della variabile *sales* rispetto alla variabile *year*

- Il numero di vendite mensili mantiene negli anni un valore minimo al di sotto di 100 ma il valore massimo incrementa da valori attorno a 300 a valori attorno a 400.
- I quartili hanno andamento crescente negli anni, con l'eccezione del secondo anno di riferimento.



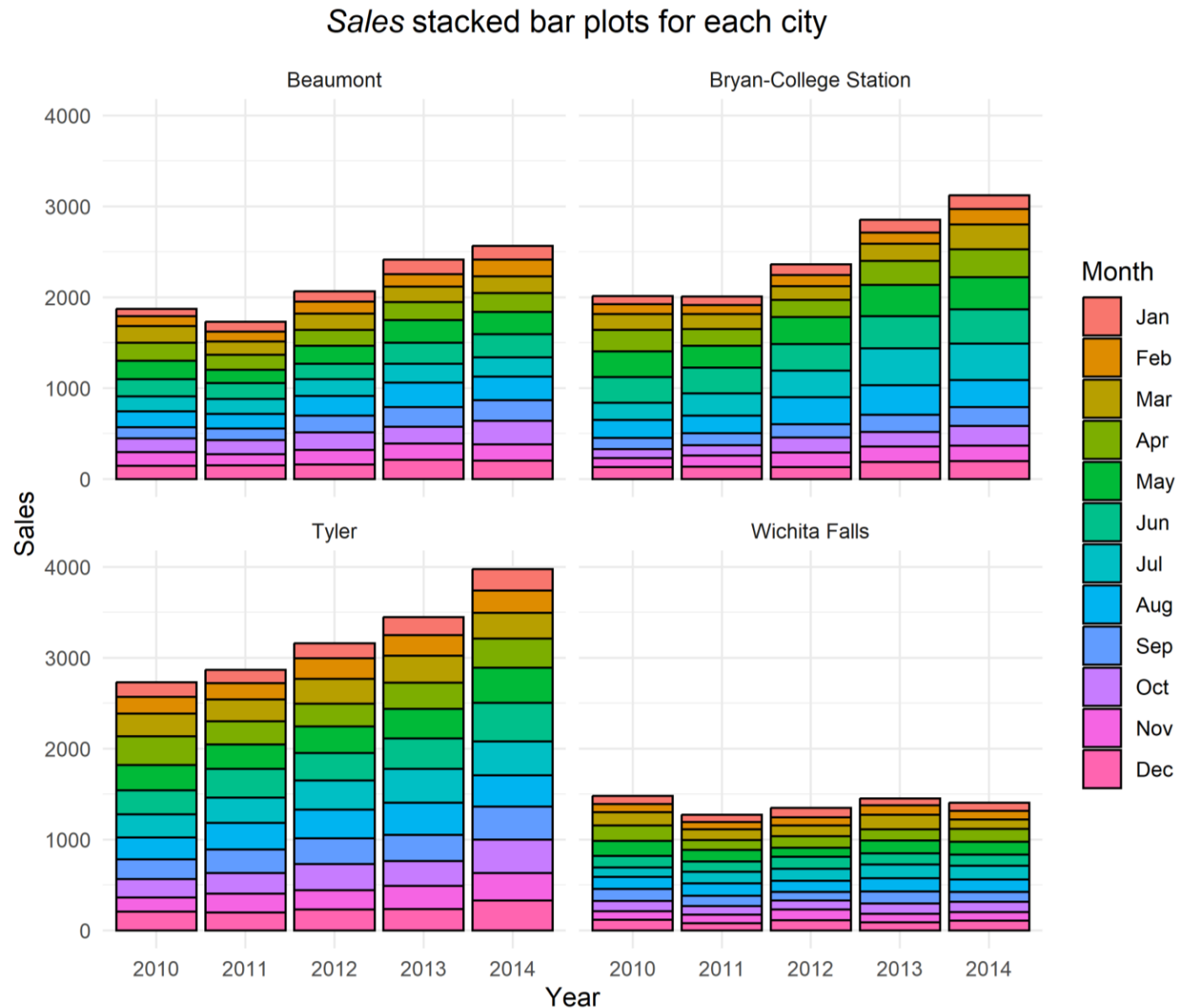
Boxplot della variabile *sales* rispetto alle variabili *city* e *year*

- Si osserva che le vendite mensili sono mediamente più basse e meno crescenti negli anni a Wichita Falls, seguita da Beaumont, Bryan-College Station e infine Tyler.
- Le vendite mensili annuali hanno una distribuzione più ampia a Bryan-College Station ma i quartili sono mediamente più alti a Tyler.



Bar plot della variabile *sales* rispetto alla variabile *year*

- Il numero di vendite cresce (tendenzialmente) negli anni in tutte le città tranne Wichita Falls, dove ha un andamento variabile.
- I valori annuali più alti si raggiungono a Tyler, seguita da Bryan-College Station, Beaumont e infine Wichita Falls.

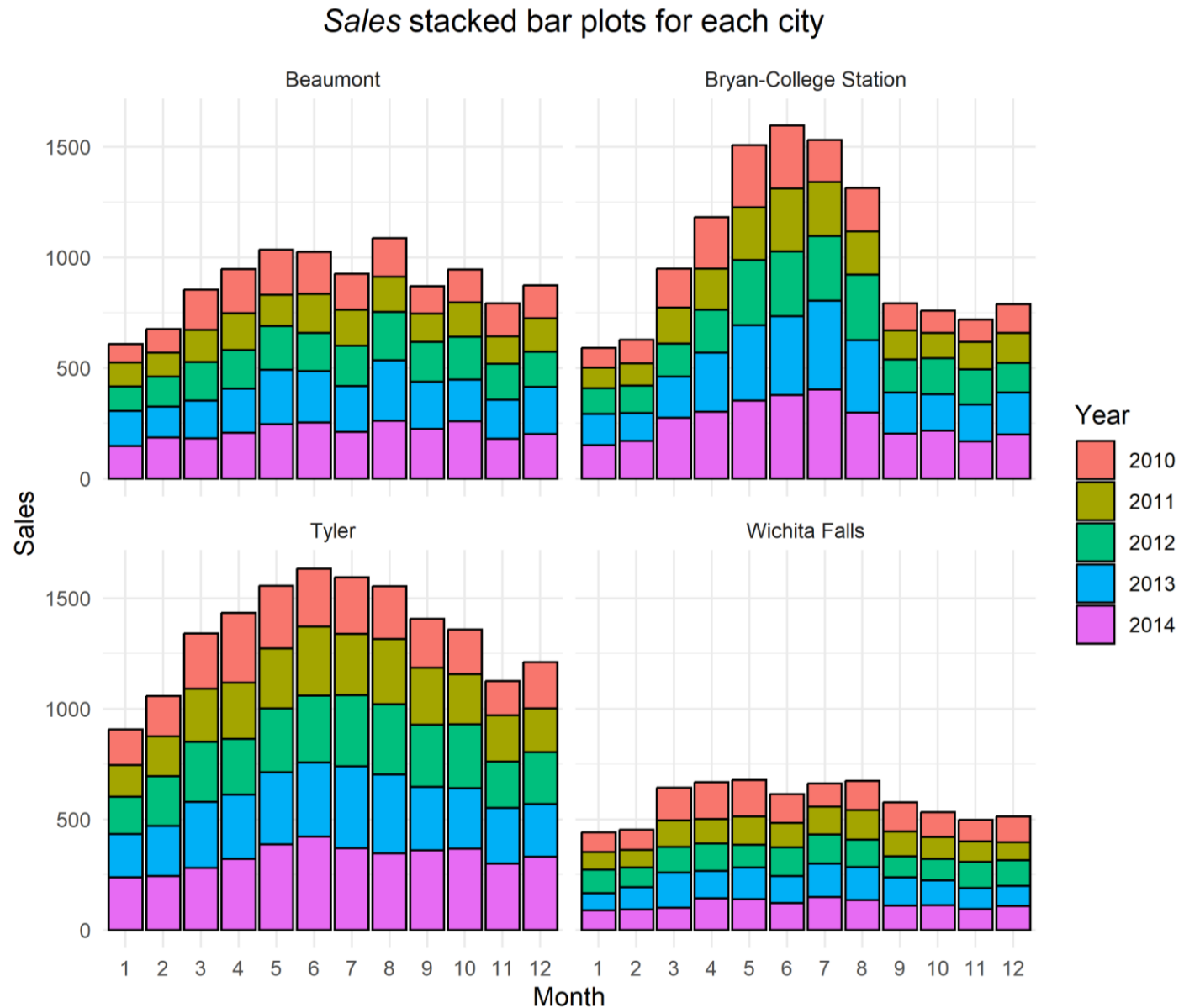


Bar plot della variabile *sales* rispetto alla variabile *month*

Considerando la totalità degli anni di osservazione:

- Gennaio e Febbraio risultano i mesi con meno vendite.
- I mesi con più vendite sono tendenzialmente quelli nella parte centrale dell'anno.

Osservando uno specifico mese, il numero di vendite è tendenzialmente crescente negli anni in tutte le città tranne Wichita Falls, dove l'andamento negli anni è più variabile.



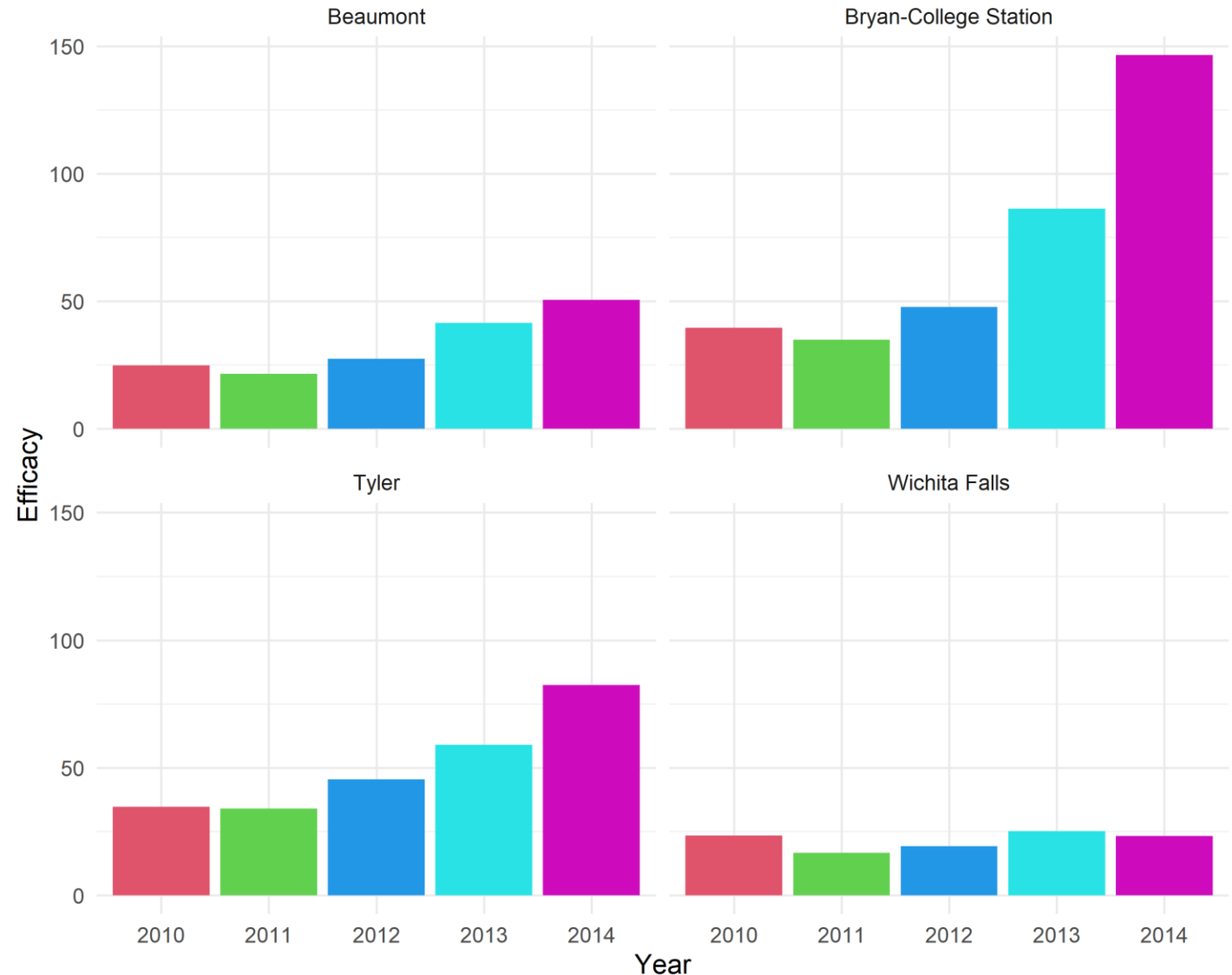
Nuove variabili: mean & efficacy

```
data$mean=volume/sales  
data$efficacy=sales*volume/(listings*months_inventory)
```

Efficacia degli annunci di vendita

- L'efficacia di vendita è stata definita come il rapporto tra il valore totale delle vendite (*volume*) e la quantità di tempo necessaria per vendere tutte le inserzioni correnti al ritmo attuale delle vendite (*months_inventry*) e vuole quindi essere rappresentativa della riuscita delle vendite sia nel passato che nel futuro.
- Essa tende ad aumentare negli anni in tutte le città tranne Wichita Falls, dove l'andamento è variabile e la crescita meno importante. L'andamento in ogni città è analogo a quello riscontrato sulla variabile *sales* (vedi slide 20), che è direttamente proporzionale alla variabile *volume*.
- La città che va meglio è Bryan-College Station, seguita da Tyler, Beaumont e infine Wichita Falls. L'ordine di "intensità" tra le diverse città è simile a quello seguito dalla variabile *sales*, con l'inversione delle città di Tyler e Bryan-College Station: sebbene la città di Tyler abbia un valore totale di vendite maggiore (vedi slide 24), in essa si riscontra una minore efficacia degli annunci di vendita, dovuta, come si osserva nelle slide 25, a un maggior tempo necessario a vendere tutte le inserzioni correnti (*months_inventry*).

Efficacy vs year bar plot for each city



Volume vs year bar plot for each city



Months of inventory vs year bar plot for each city



Probabilità

city	Beaumont	Bryan-College Station	Tyler	Wichita Falls
	60	60	60	60

month	1	2	3	4	5	6	7	8	9	10	11	12
	20	20	20	20	20	20	20	20	20	20	20	20

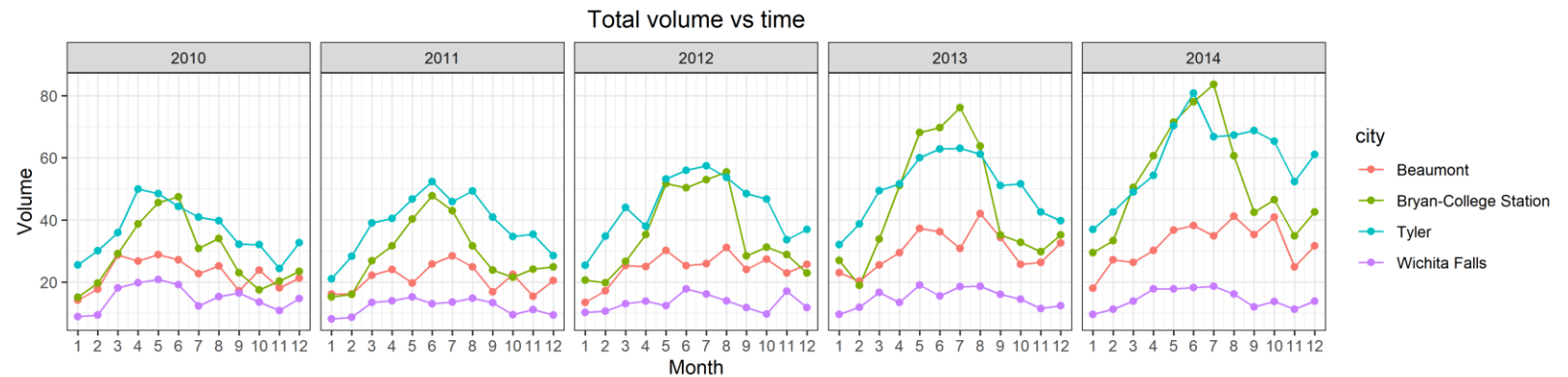
	month	1	2	3	4	5	6	7	8	9	10	11	12
year	2010	4	4	4	4	4	4	4	4	4	4	4	4
2011	4	4	4	4	4	4	4	4	4	4	4	4	4
2012	4	4	4	4	4	4	4	4	4	4	4	4	4
2013	4	4	4	4	4	4	4	4	4	4	4	4	4
2014	4	4	4	4	4	4	4	4	4	4	4	4	4

Dal numero di occorrenze per le diverse modalità delle variabili *city*, *month* e *year* risulta che:

- La probabilità di prendere un'osservazione in cui *city* sia uguale a Beaumont è di $60/240=0.25$
- La probabilità di prendere un'osservazione in cui *month* sia uguale a Luglio (7) è di $20/240=0.083$
- La probabilità di prendere un'osservazione in cui *month* sia uguale a Dicembre (12) e *year* sia uguale a 2012 è di $4/240=0.016$

Line chart della variabile *volume*

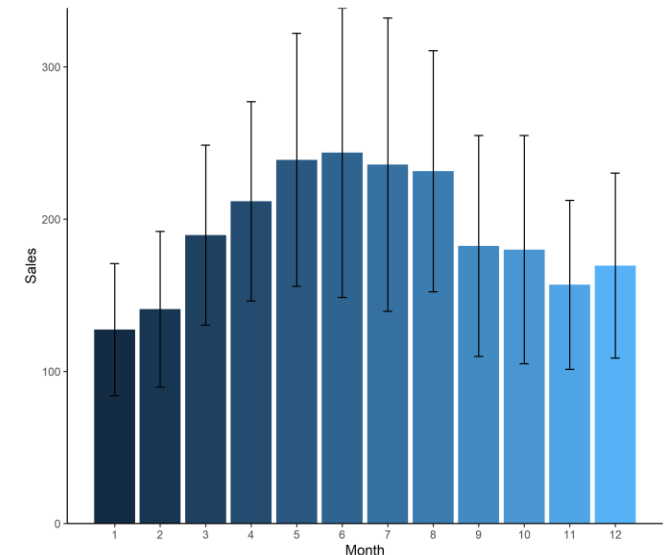
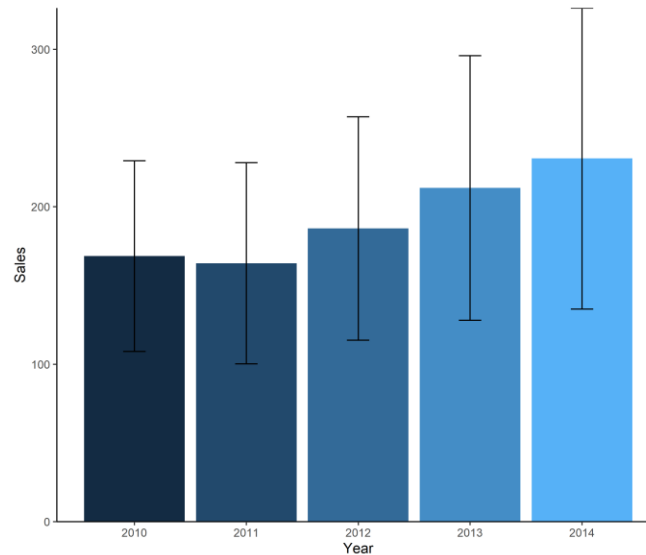
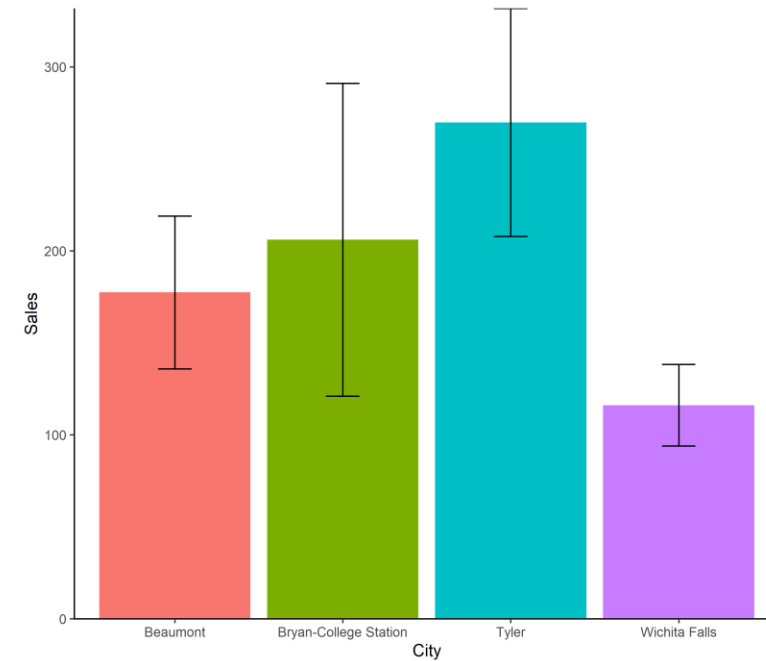
- Il valore totale delle vendite ha un andamento variabile durante ogni anno, indipendentemente dalla città, con valori meno variabili a Wichita Falls e picchi raggiunti durante la parte centrale dell'anno a Bryan-College Station e Tyler.
- Dato un anno specifico, le curve si intersecano poche volte, indicando, indipendentemente dall'anno e dal mese di riferimento, un ordine di "intensità" tra le città mediamente costante, partendo da valori più bassi a Wichita Falls, seguita da Beaumont, Bryan-College Station e Tyler, come accade per la variabile *sales*.



Media e deviazione standard della variabile *sales*

I grafici mostrano vendite con media:

- Crescente secondo l'ordine di Wichita Falls, Beaumont, Bryan-College Station e Tyler (come osservato nella slide 17)
- Crescente negli anni con eccezione del 2011 (come osservato nella slide 18)
- Crescente fino a giugno e con picchi nei mesi primaverili ed estivi (come osservato nella slide 21)



Media e deviazione standard della variabile *months_inventory*

I grafici mostrano un months of inventory medio:

- Più basso nelle città di Wichita Falls e Bryan-College Station e maggiore a Beaumont e Tyler
- Con andamento decrescente negli anni ad eccezione del 2011
- Con andamento crescente fino ad aprile e picchi nei mesi primaverili ed estivi

