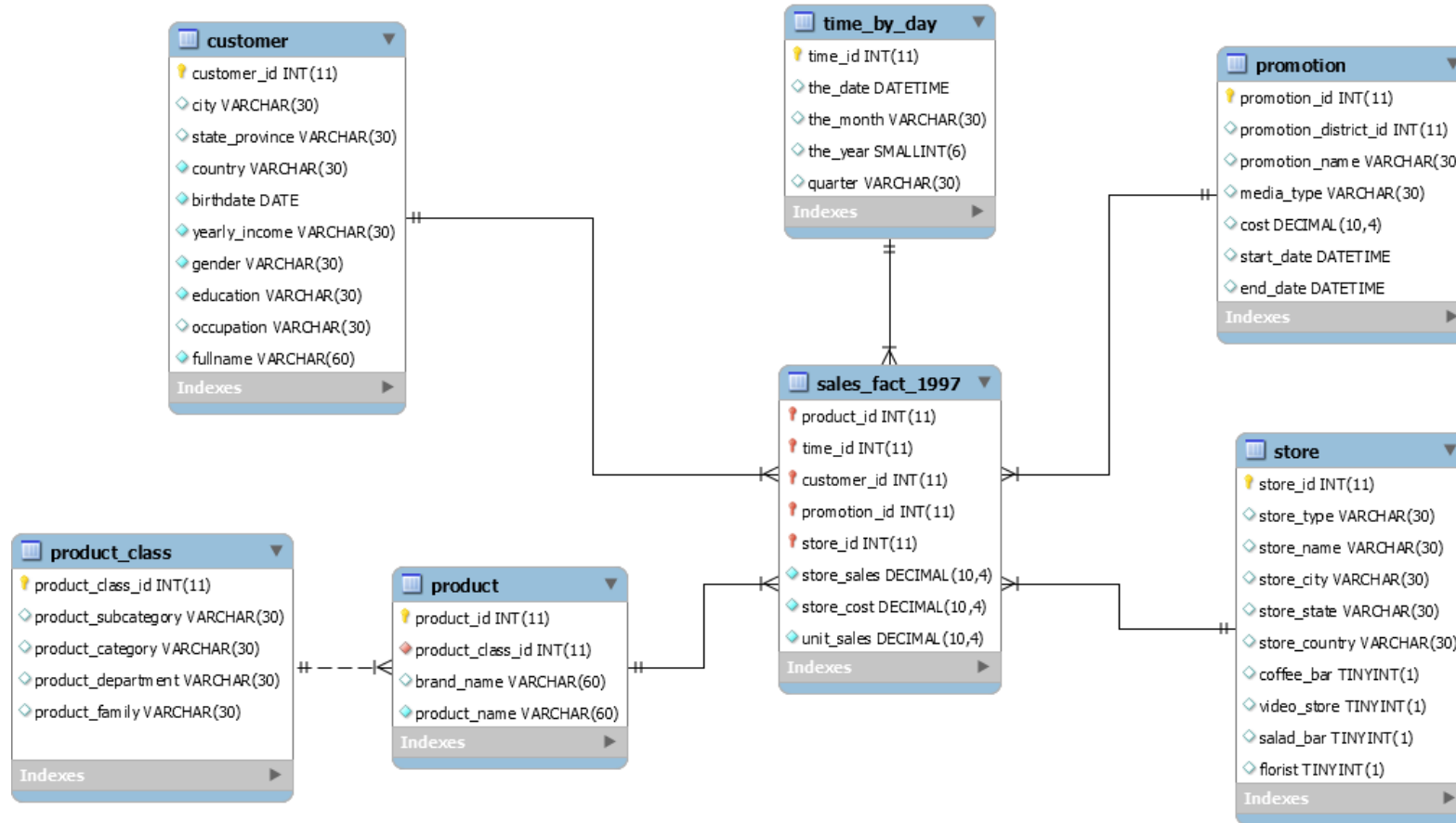


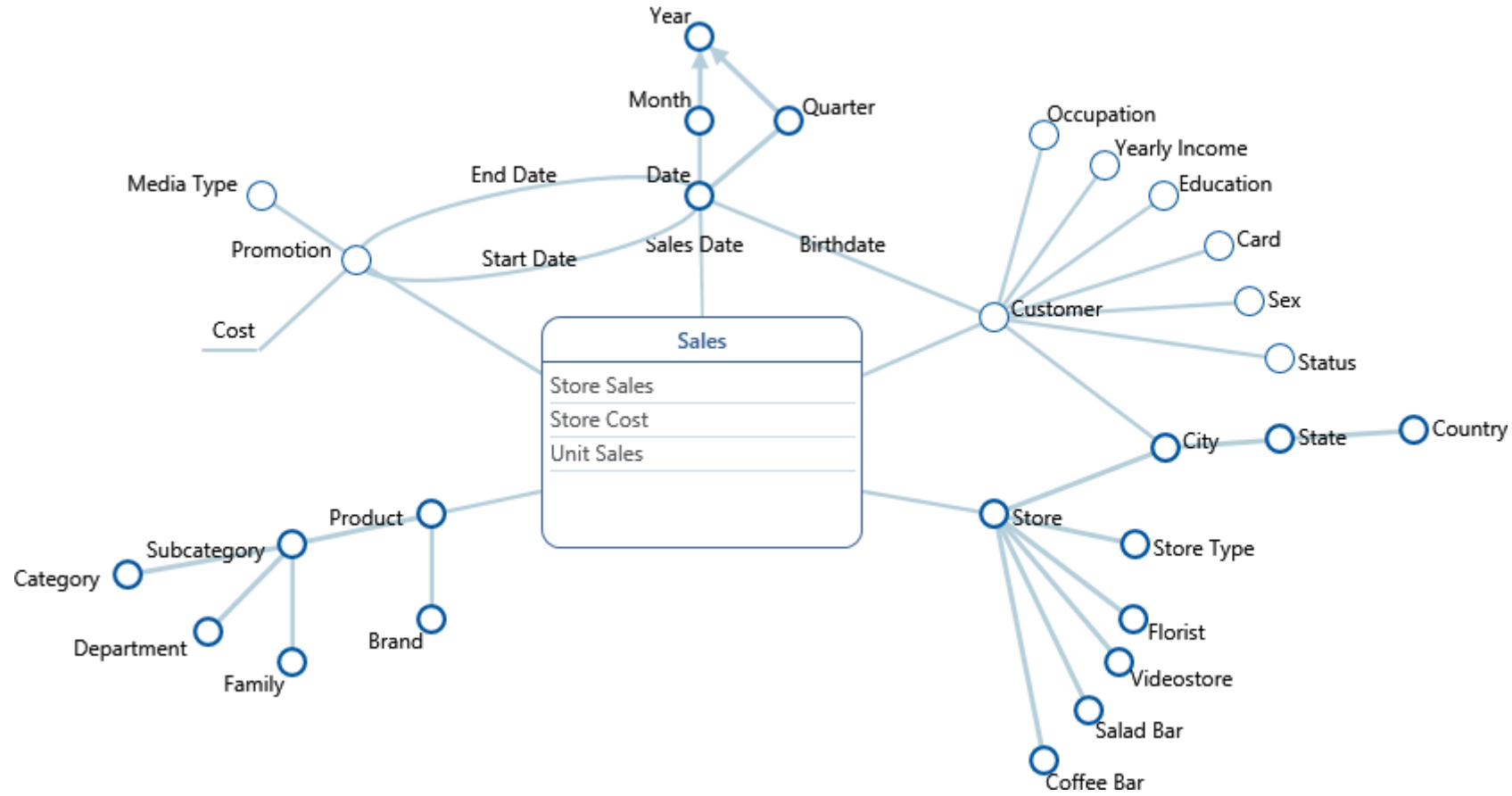
OLAP and visualization

Hands on Tableau

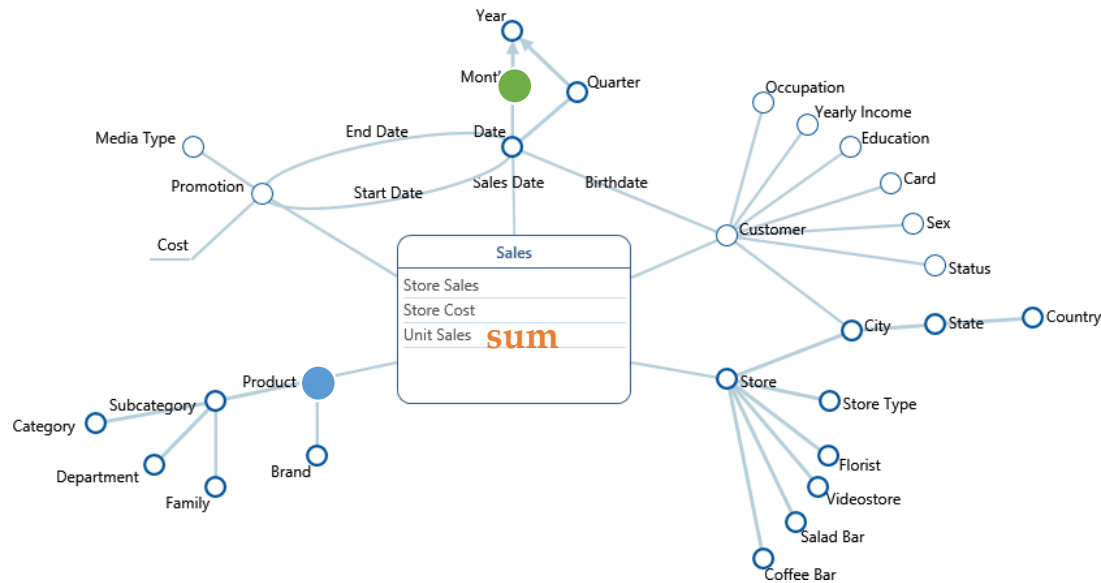
Sales (Foodmart) – Database Structure



Sales (Foodmart) – DFM

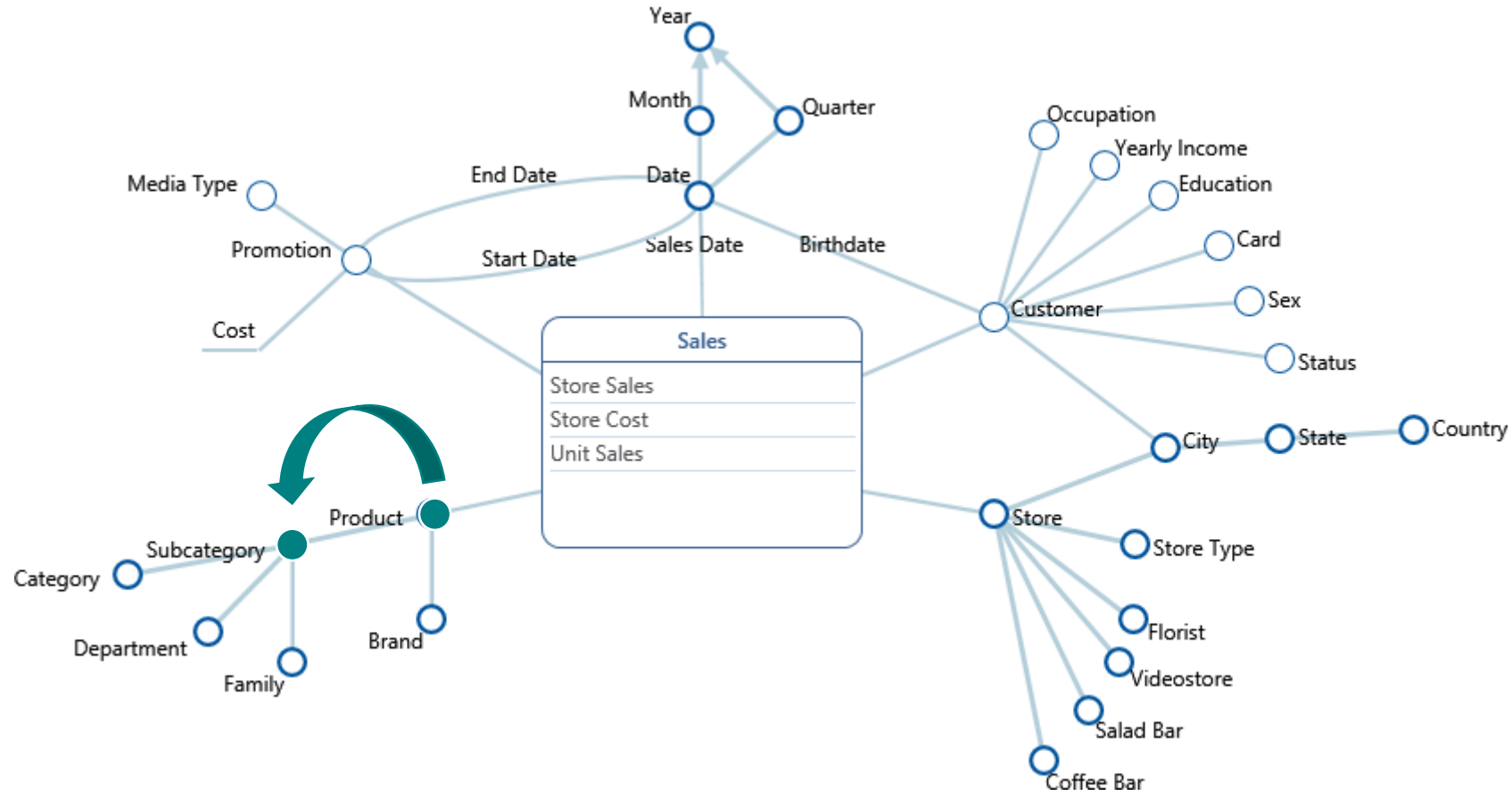


Operatori OLAP: Cube query

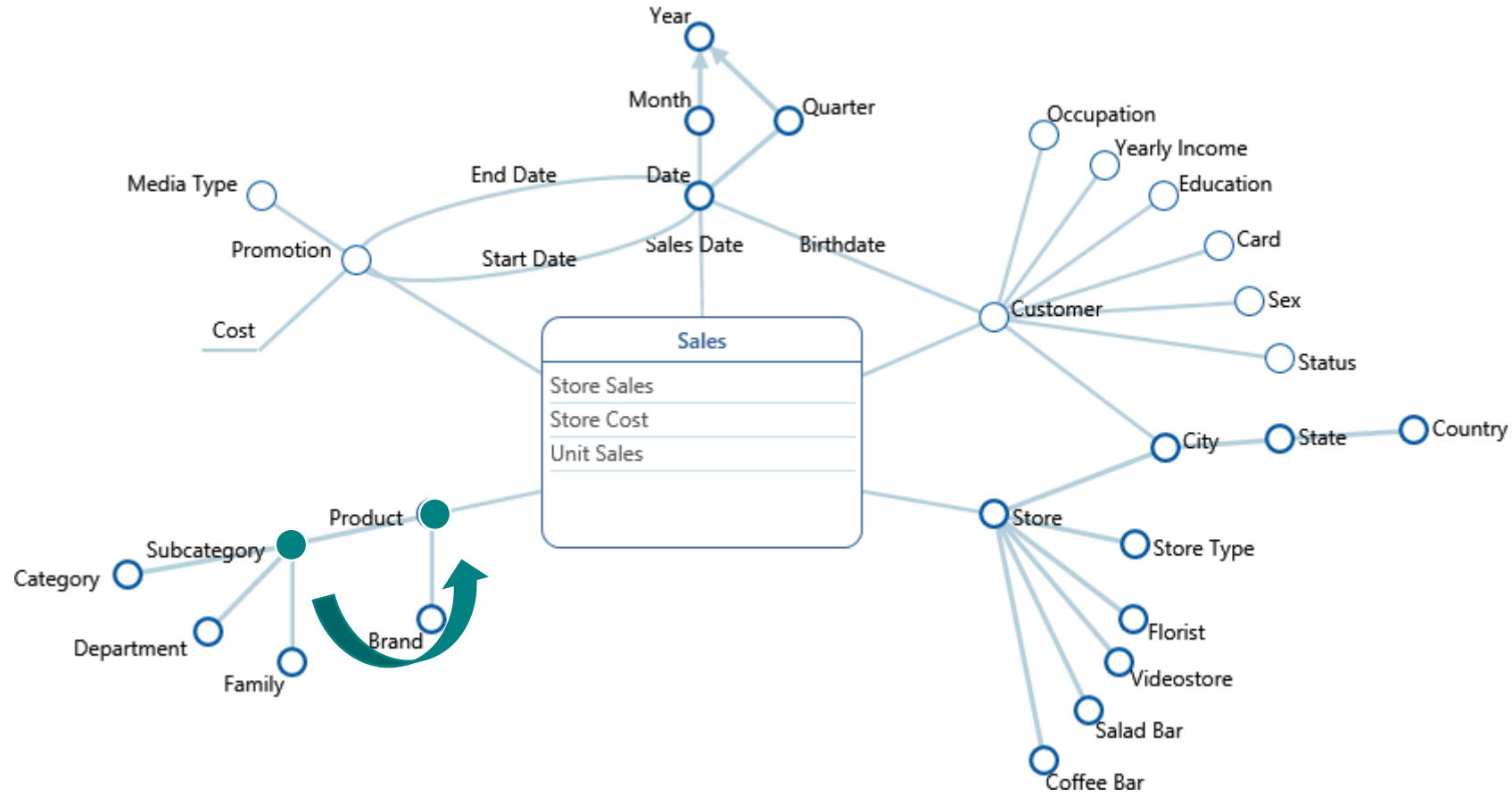


Get the **sum** of sales
by **month** and **product**

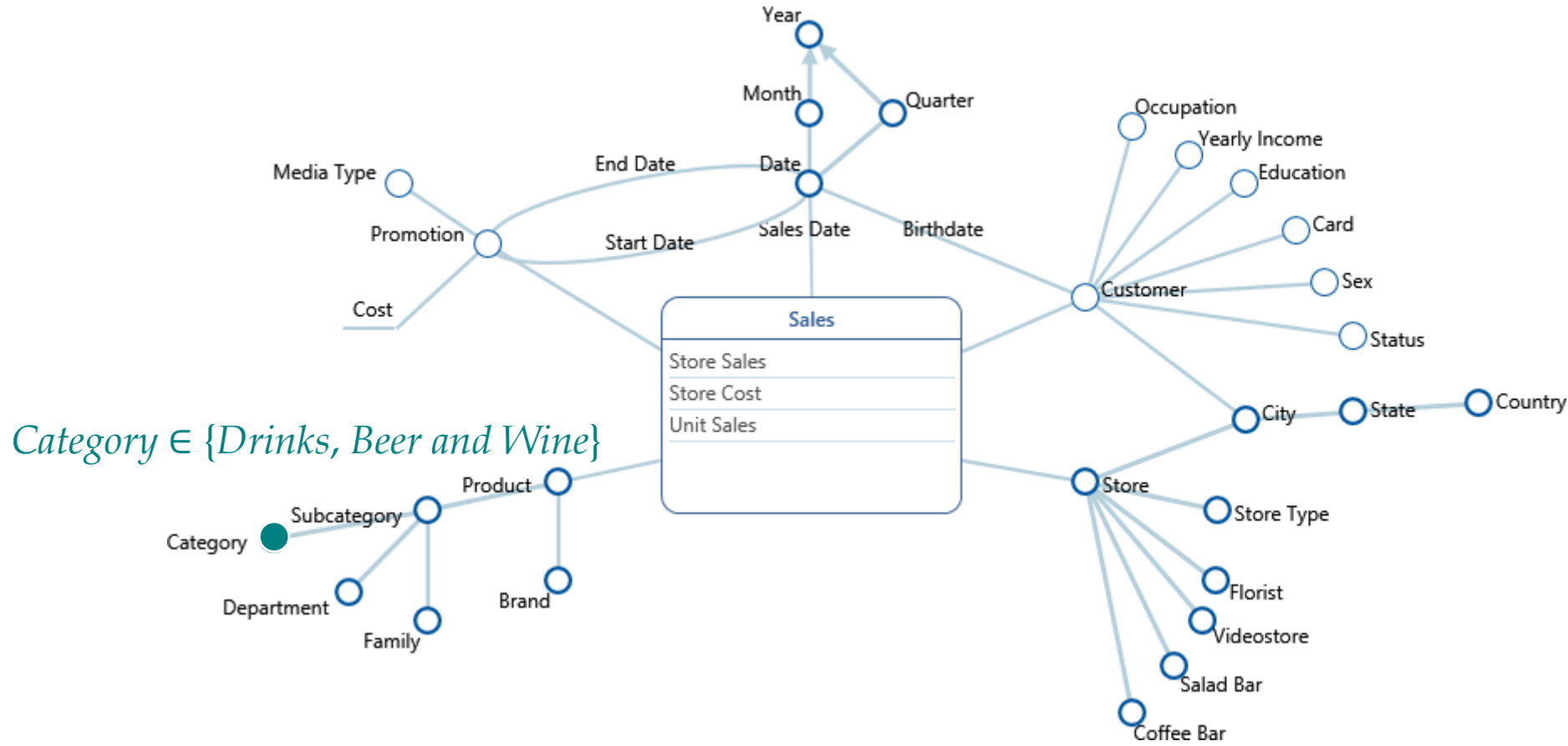
Operatori OLAP: Roll-Up



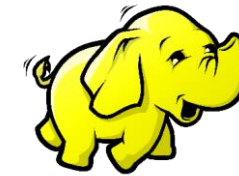
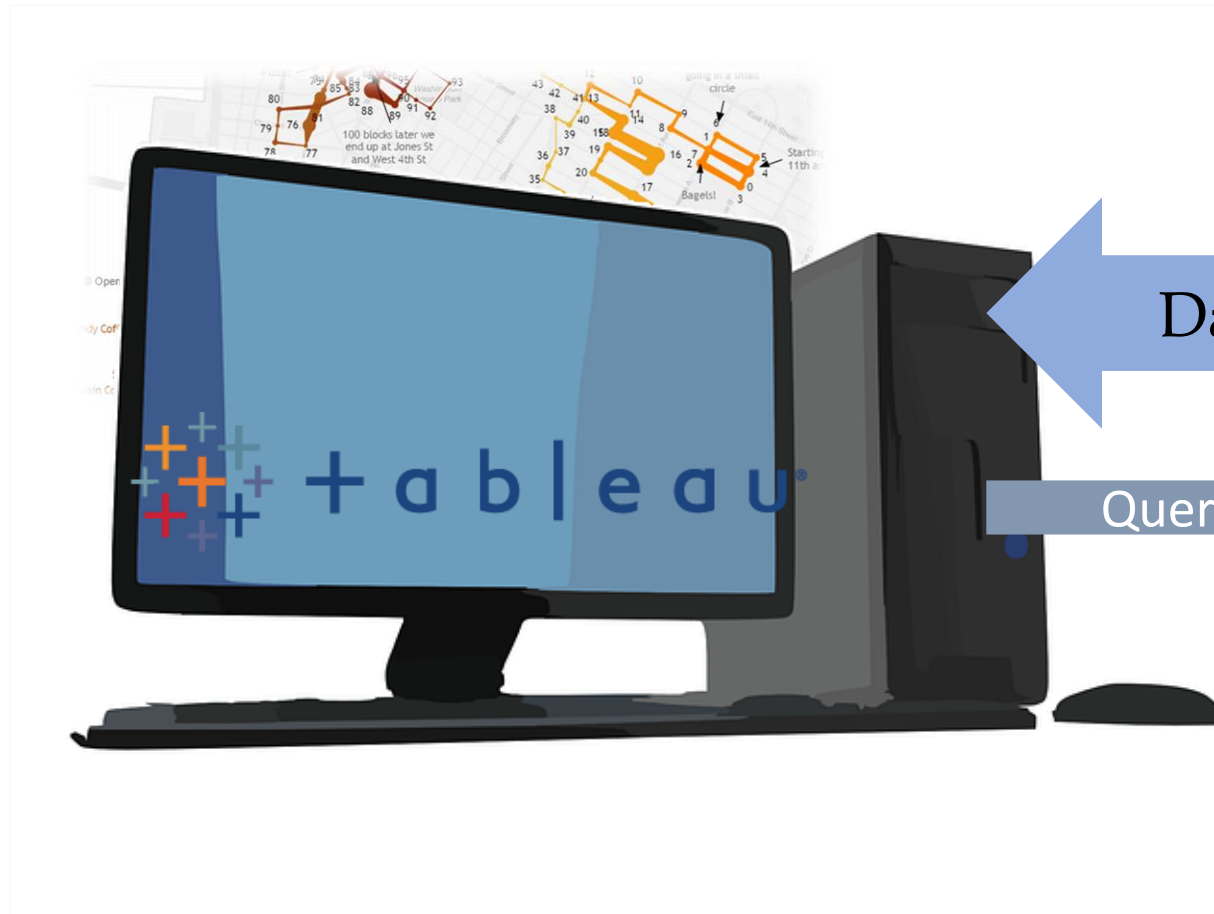
Operatori OLAP: Drill-Down



Operatori OLAP: Slice & Dice



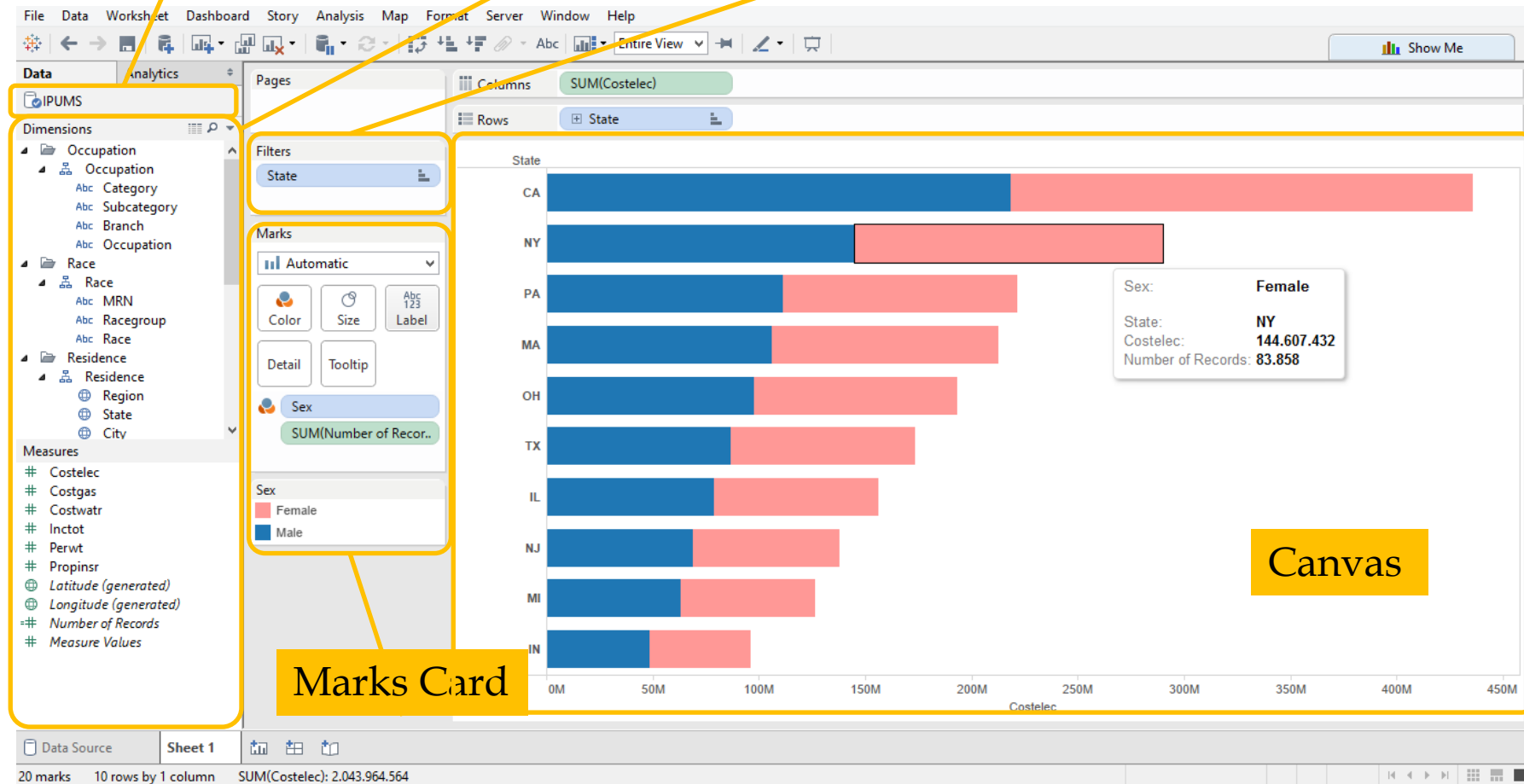
Architettura

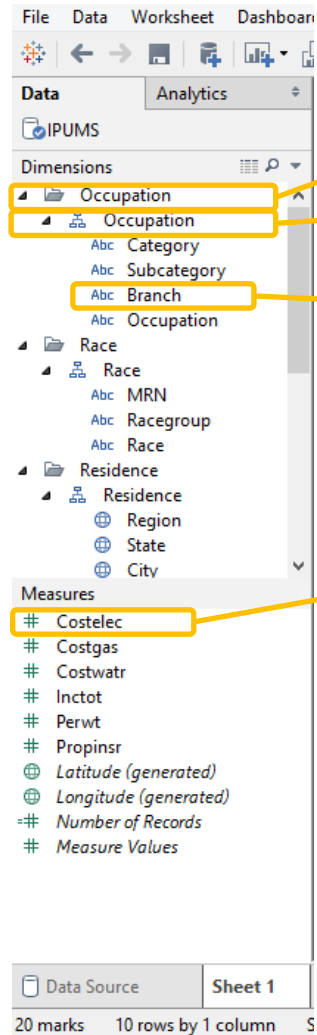


Sorgenti dati

Dimensioni e misure

Filtri





Dimensione

Gerarchia

Attributo dimensionale

Misura

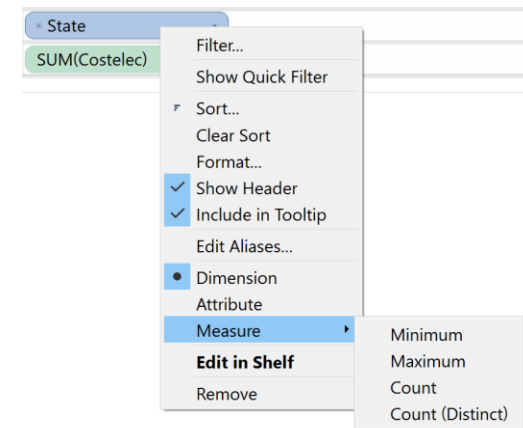
Per convenzione, le **cartelle** vengono usate per rappresentare **dimensioni**; in generale è solo un modo per raggruppare elementi

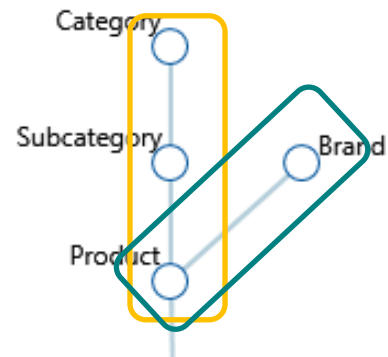
Dimensione VS Misura

In Tableau le definizioni di *dimensione* e *misura* sono più lasche rispetto a quelle tradizionalmente utilizzate in letteratura e ogni campo può essere utilizzato sia come dimensione che come misura

In generale è comunque utile dare una classificazione iniziale ai campi

- Una dimensione è un qualunque campo **indipendente** (e.g., *città*)
- Una misura è un qualunque campo i cui valori sono **funzione di altri campi** (e.g., *profitto vendite*)

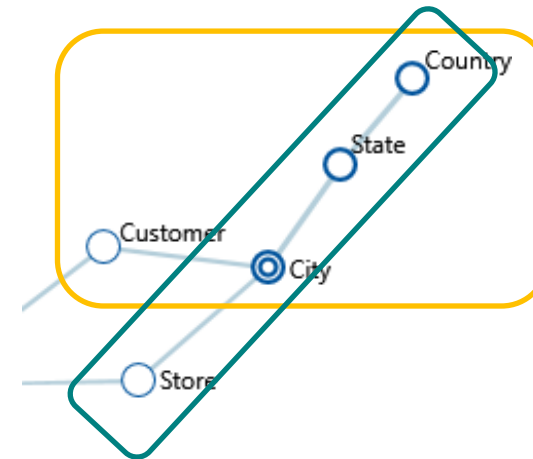




Ogni path dalle foglie alla radice diventa una gerarchia separata e gli attributi comuni vengono duplicati

Le gerarchie condivise vengono duplicate

S. Country S. State S. City



Green VS Blue

In Tableau, il colore verde è associato a campi **continui** mentre il colore blu a quelli **discreti**



Spesso (ma non sempre) le misure sono campi continui, mentre le dimensioni sono campi discreti

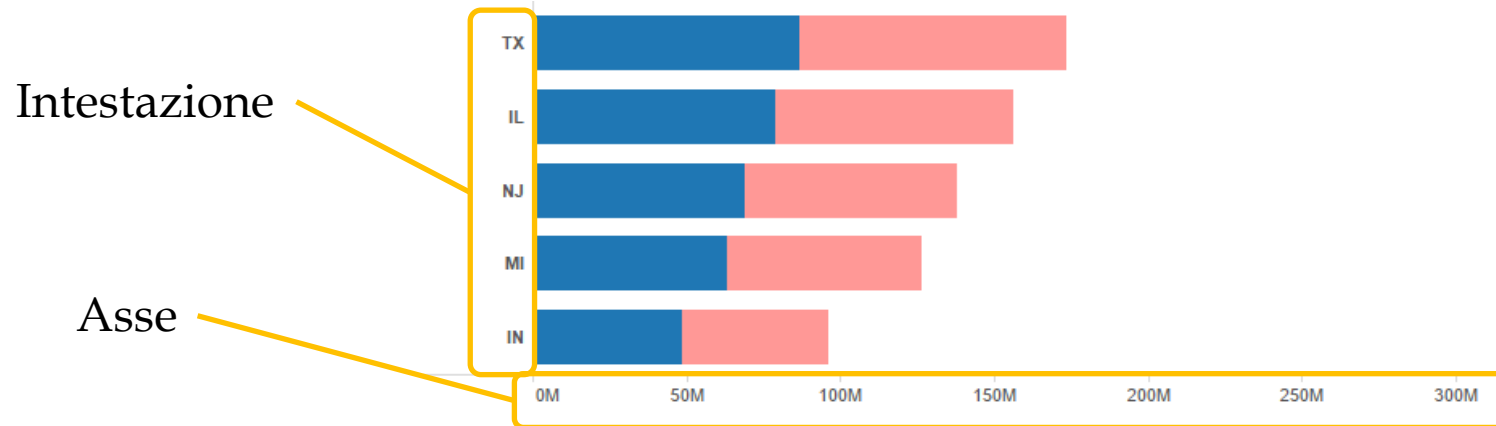
Campi continui e discreti producono effetti diversi

- Quando vengono impostati in righe e colonne
- Quando vengono utilizzati in un filtro
- Quando vengono associati a colori

Green VS Blue (Righe e Colonne)

Quando vengono assegnati a righe e colonne

- Un campo *discreto* produce un'*intestazione*
- Un campo *continuo* produce un *asse*



Green VS Blue (Filtri)

Con un campo continuo è possibile specificare dei **range**

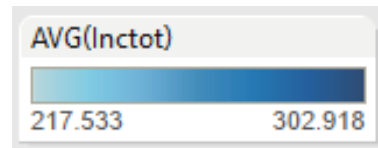
- Sui valori al livello più **dettagliato**
- Oppure su particolari **aggregazioni**

#	All values
#	Sum
#	Average
#	Median
#	Count
#	Count (Distinct)
#	Minimum
#	Maximum
#	Standard deviation
#	Standard deviation (Population)
#	Variance
#	Variance (Population)

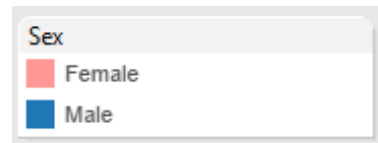
Con un campo discreto è possibile anche selezionare valori specifici (i.e., uno ad uno)

Green VS Blue (Colori)

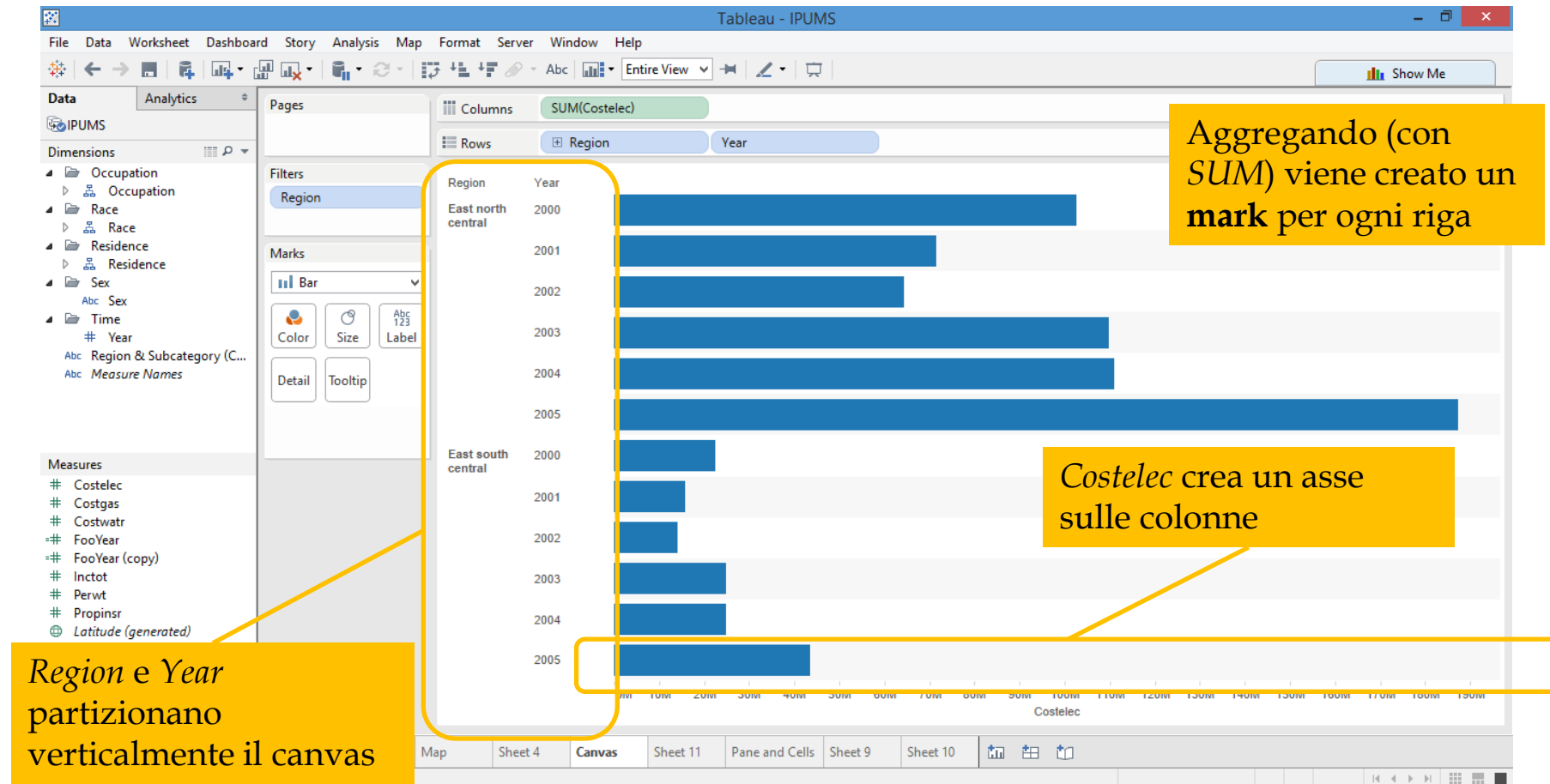
Ad un campo continuo viene associato un insieme di colori **sequenziali** (e.g. gradazioni più chiare per valori bassi e gradazioni più scure per valori alti)



Con un campo discreto è possibile assegnare un colore diverso (non necessariamente correlato agli altri) per ogni valore distinto



Canvas

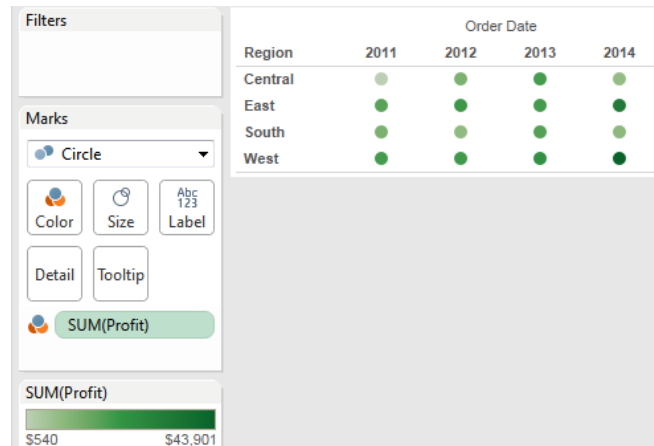


Mark

I dati vengono visualizzati all'interno del canvas tramite *mark*

Esistono diversi tipi di mark (*bar*, *line*, *text*, etc.)

- Ogni mark possiede diverse *proprietà* (*colour*, *size*, *label*, etc.)
- A ciascuna proprietà può essere associato un campo con effetti diversi in base alla proprietà e al tipo di campo (continuo o discreto)

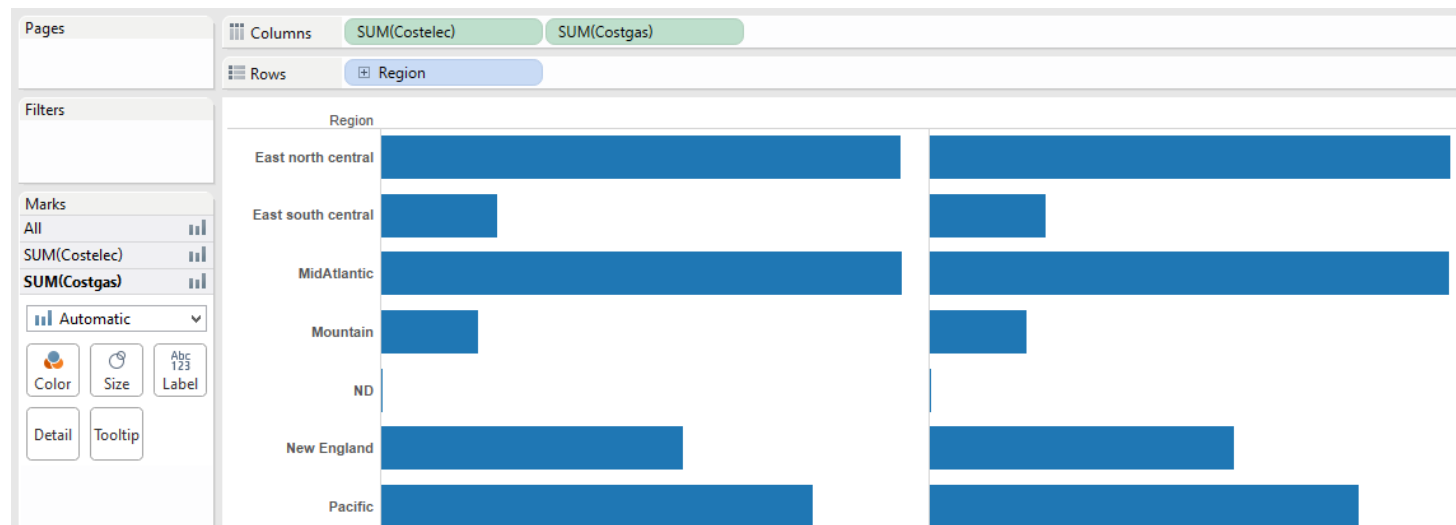


Mark

Una cell può contenere da zero a più mark

- Ad esempio nel caso in cui vengano visualizzate più misure

In uno stesso canvas possono essere presenti più tipi di mark

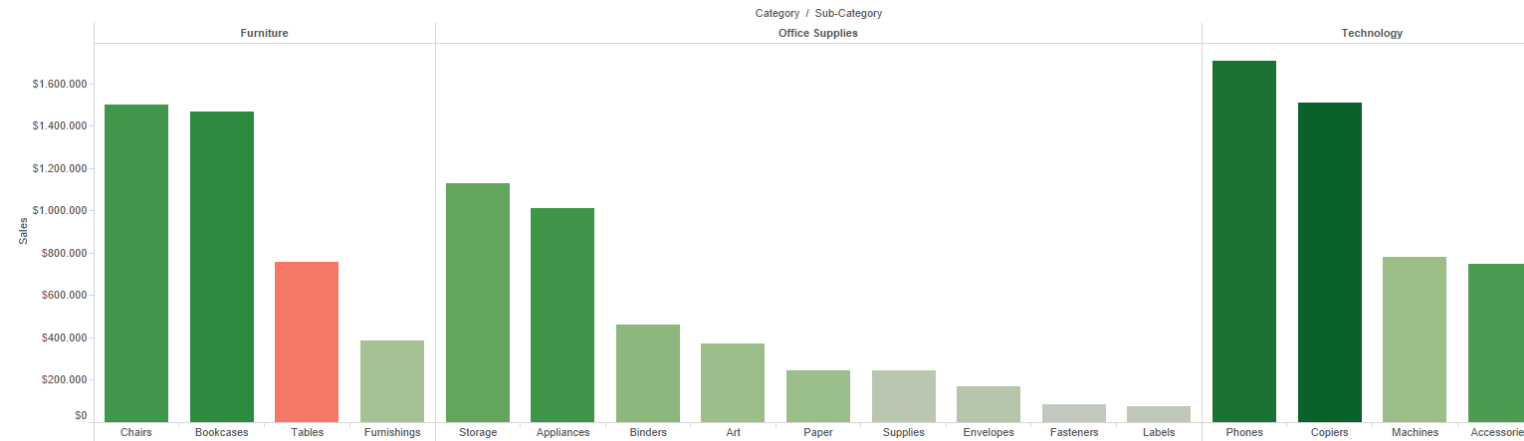


Ordinamento

Esistono tre tipologie di ordinamento

- *Manual*: l'ordinamento è fissato manualmente dall'utente
- *Computed*: l'ordinamento si basa su un calcolo (eg. la somma di una certa misura)
- *Data Source Order*: l'ordinamento è lo stesso della sorgente dati

Gli ordinamenti rispettano le gerarchie (ci sono workaround)

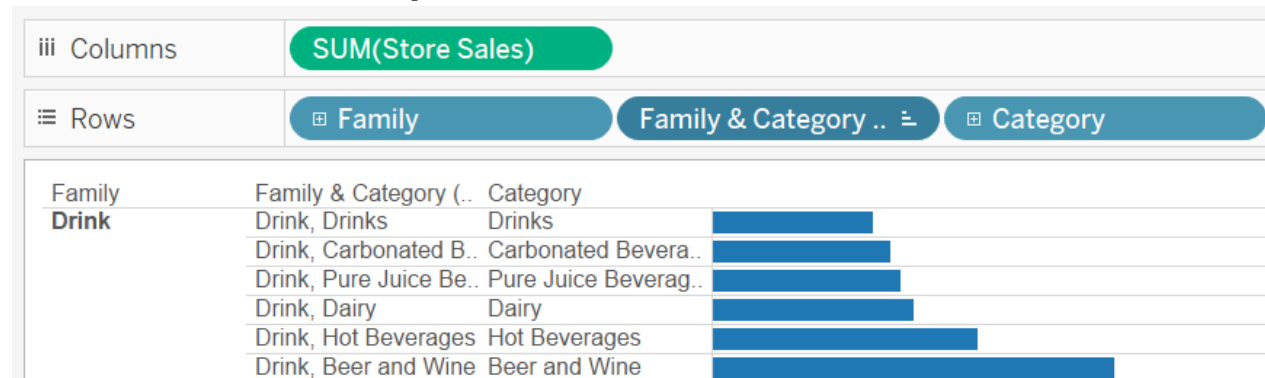


Ordinamento

Un comportamento inaspettato si presenta quando si cerca di ordinare un campo a destra di un altro che non lo determina funzionalmente



Workaround: utilizzare un campo combinato



View Data

Tramite l'opzione *View Data* è possibile visualizzare l'insieme record (i.e., i dati a granularità più fine) utilizzati per calcolare un determinato mark

View Data può essere considerato come una versione light dell'operazione *Drill Through*

Particolarmente utile per test e debug quando si creano visualizzazioni complesse

View Data (2)

The screenshot shows a map application interface. On the left, a sidebar contains a search icon, zoom in (+), zoom out (-), a pin icon, and a play icon. The main map area displays the United States, with California highlighted in blue. A context menu is open over California, listing various actions: Select All, View Data... (highlighted), Copy, Format..., Edit Locations..., Mark Label, Annotate, Trend Lines, Forecast, Drop Lines, Hide Map Search, Hide View Toolbar, Keep Only (checked), Exclude, Group, and Create Set....

In the foreground, a 'View Data' window is open, showing a table with 75,000 rows. The window has a blue title bar and a toolbar with 'Copy' and 'Export All' buttons. The table has three columns: City, MRN, and Occupation. The data is as follows:

City	MRN	Occupation
Los Angeles	1	First-Line Supervisors/Managers of Construction Trades and Extraction Workers
Los Angeles	1	Painters, Construction and Maintenance
Los Angeles	1	Carpenters
Los Angeles	1	Construction Laborers
Los Angeles	1	Carpenters
Pomona	1	Electricians
Riverside	1	Electricians
San Diego	1	Carpet, Floor, and Tile Installers and Finishers
Los Angeles	1	Electricians
Los Angeles	1	Carpenters

At the bottom of the window, there are tabs for 'Summary' and 'Underlying', and a status bar indicating '75,000 rows'.

Show Me

La palette *Show Me* contiene scorciatoie per produrre visualizzazioni di tipi differenti a partire da un insieme di dimensioni e misure

Per poter utilizzare una visualizzazione tramite Show Me è necessario rispettare determinati *requisiti* che variano di caso in caso (e.g. per uno scatter plot sono necessari campi continui)

Alcuni tipi di visualizzazioni sono poco intuitive da costruire manualmente (e.g. mappe e box-plot)



For **scatter plots** try
0 or more dimensions
2 to 4 measures

Esercizio 1

Visualizzare tramite un grafico a barre la somma delle *Store Sales* per ogni *S. State*

- Qual è lo stato con le vendite più elevate?

Effettuare un drill-down per visualizzare le vendite a livello di *S. City*

- Esistono città con vendite molto inferiori rispetto alle altre?

Quanti sono i negozi (*Store*) presenti in ogni *S. State*?

E in ogni *S. City*?

- **Tip:** Utilizzare l'aggregazione *COUNTD*
- È possibile imputare le basse vendite in alcune città al numero di negozi?

Visualizzare le vendite a livello di *S. City* (come fatto in precedenza) e associare alla proprietà *color* numero di negozi distinti (*COUNTD(store)*)

Esercizio 2

Data l'ultima visualizzazione creata in Esercizio 1, associare il campo *S. Type* alla proprietà *color*

- Quale pattern interessante è possibile notare?

Visualizzare le *vendite (SUM)* per ogni *S. Type*

- Quale discrepanza è possibile notare rispetto al grafico precedente?

Associare il *numero di negozi (Store)* alla proprietà *color* e alla proprietà *label*

- Da cosa è causata la discrepanza tra le due visualizzazioni precedenti?

Esercizio 3

Visualizzare tramite un grafico a linee l'andamento **mensile** delle **vendite**

- Quale pattern è presente?

Dividere il grafico precedente per **S. State** (un asse per ogni stato)

- Il pattern precedente è presente in ogni stato?
- **Tip:** di default gli assi hanno tutti lo stesso range: su un asse qualsiasi, click destro > *Edit Axis* > Selezionare *Independent axis...*

Dato il grafico precedente, visualizzare quanto impattano le varie **Family** sul totale delle vendite mantenendo la visualizzazione del trend mensile

- Quale può essere una buona visualizzazione?
- **Tip:** associare ogni **Family** ad una proprietà dei mark ed eventualmente cambiare tipologia di mark
- **Tip:** è possibile cambiare il tipo di mark dal menu a tendina nel pannello *Marks*