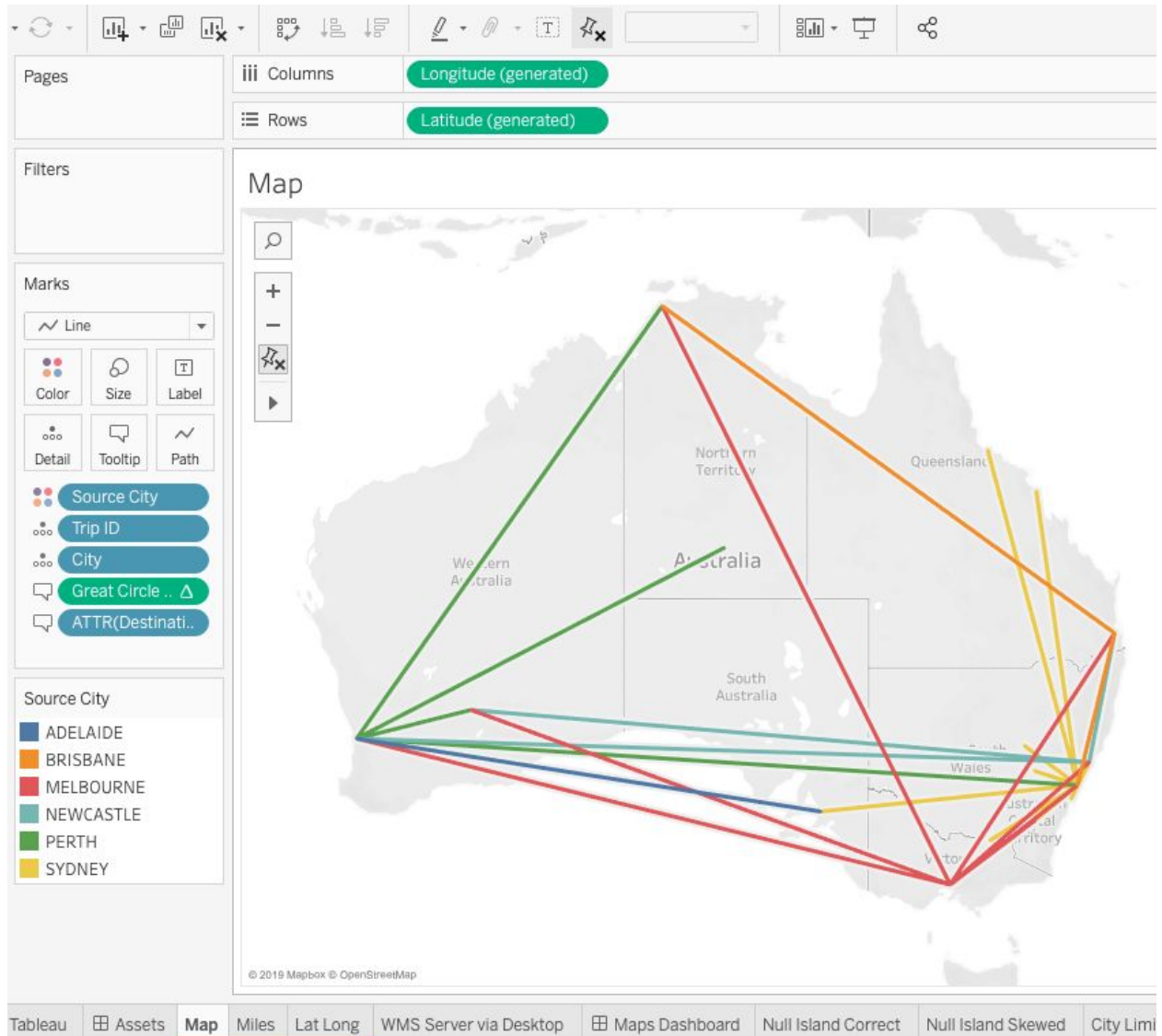


Assignment 3: Tableau Maps, Dashboards, and Animation

Team member: Chuyue Wu, Menghong Han, Xiaofan Sun, Yifan He

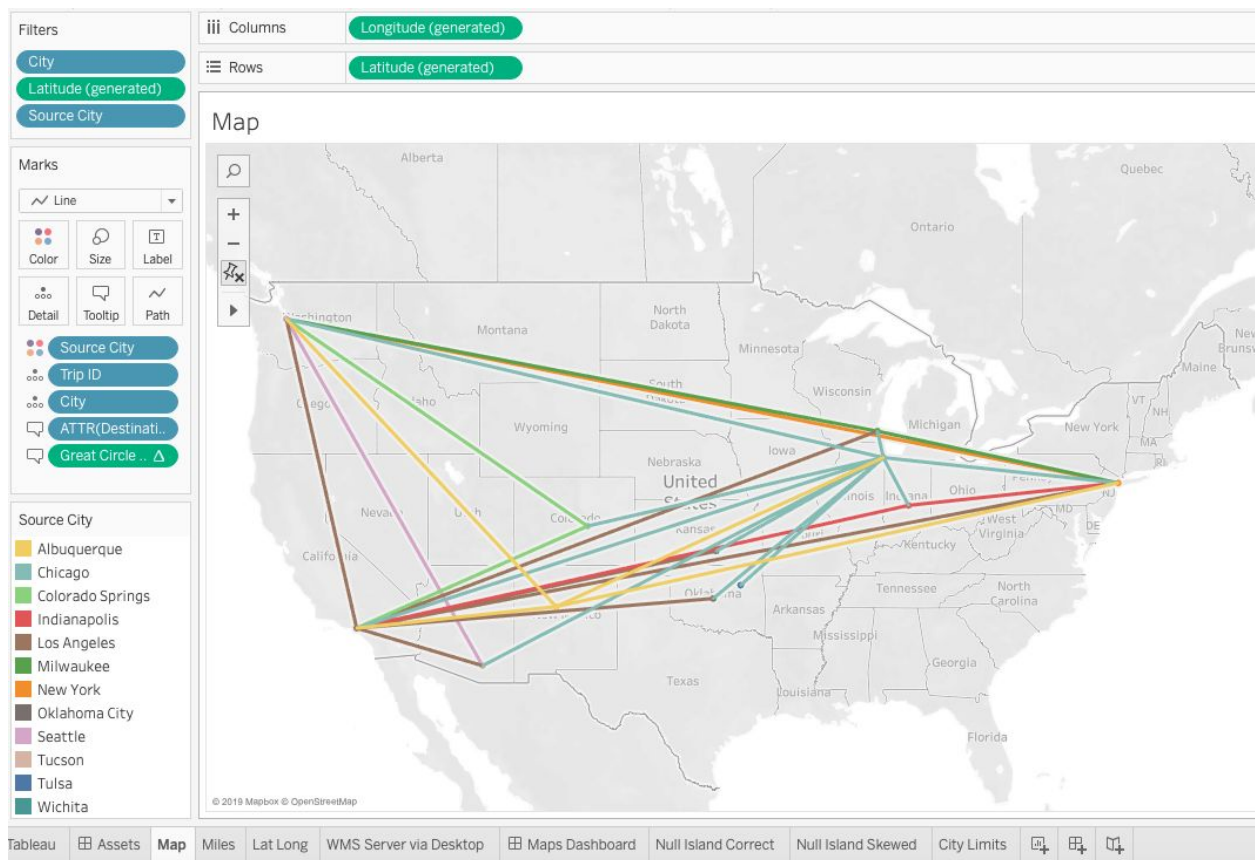
A. Chapter 9, page 272, displaying routes and calculating distances



(Replication: Australia)

Brief Introduction of Replication:

This is our screenshots from replication. It shows six main source cities in Australia and flight routes between these source cities and destination cities.



(Application: United States)

Brief Introduction of Application:

This is our screenshots from application. The new dataset is from the openflights website (<https://openflights.org/data.html>). We download *airports.dat* and *routes.dat*, merge them by airport ID, select US data, delete unnecessary columns, and melt them in the same structure as the replication data does. Then, we write them out of R in a csv file, import it into Tableau dataset, and repeat the plot process in Tableau again. To make the plot more clear, I filter 50 largest US cities by population but only 12 cities have location information in Tableau(latitude generated). Compared to dataset of Australia which has only 6 source cities, we believe 12 Source cities are enough.

Analogous Features:

Firstly, after clearing data in R, two datasets have the same structure. That's to say, two datasets both have Trip ID, city, Dest/Orig, so we can repeat the plot process easily in application.

Secondly, when plotting our own data, we didn't use all city information in the dataset, but select some important cities in the US. In replication, we also only plot 6 main cities in Australia. It's a good way to focus on big cities in two countries and make the plot more clearly.

Insight:

From the replication plot in Australia, we could find that source cities in the east coast of Australia has more flight routes such as Sydney and Melbourne. Some cities in the west coast of Australia and north coast of Australia are popular destination cities such as Perth and Darwin, but they are not as busy as Sydney and Melbourne. That makes sense because harbour cities have beautiful scenery, warm climate and convenient import and export system, which could make both people increase and economy develop.

From the application plot in the US, we could find that in the US, there are also some popular cities along the coasts such as New York City, Los Angeles and Seattle. However, due to the extreme long distance from the west coast to the east coast in the US, there are something different from Australia. We could find from the plot that the busiest source city that has most flight routes is Chicago. Chicago is located in the middle of the US, which has been a famous north-south transportation hub in the United States all the time. From the plot, we could tell that Chicago still plays an important role in today's air transportation system.

B. Chapter 10, page 308, creating a template

Screenshot for template

Template



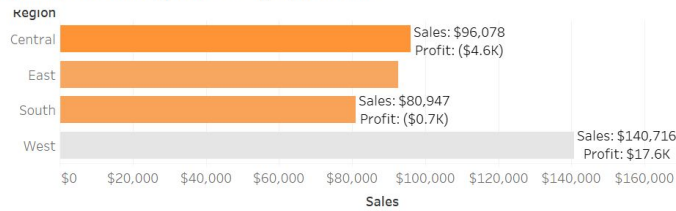
C.

For template: We use template completed in B and adjust the color of lower bar

Screenshots for replication

Superstore Loss Leaders

Loss Leader Regions - Orange represents loss



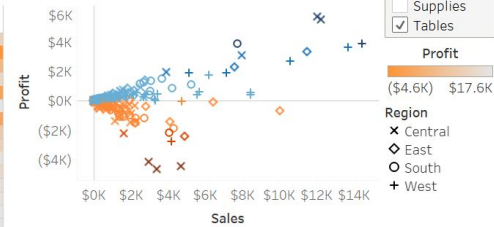
Loss Leader Cities

Orange represents loss. Blue represents profit.

Sta..	City	Avg Dsc..	Discount	Profit	Sales
Texas	Dallas	37.02%	69%	-\$1,984	\$3,507
	Houston	73%		-\$7,520	\$8,148
	San Anto..	64%		-\$4,455	\$4,710
Illinois	Aurora	39.00%	63%	-\$1,475	\$1,983
	Chicago		76%	-\$6,358	\$5,372
Pennsy..	Lancaster	32.86%	70%	-\$106	\$132
	Philadelp..		66%	-\$5,974	\$12,604
Arizona	Phoenix	30.36%	66%	-\$2,779	\$4,183
North ..	Burlington	28.35%	50%	-\$2,159	\$4,045
	Jacksonvi..		67%	-\$474	\$1,177
Florida	Jacksonvi..	29.93%	64%	-\$1,186	\$2,231
Colora..	Aurora	31.65%	67%	-\$576	\$876
Ohio	Lancaster	32.49%	65%	-\$434	\$705
Iowa	Burlington	0.00%	0%	\$3	\$5
Califor..	Lancaster	7.28%	20%	\$41	\$118

Loss Leader Zip

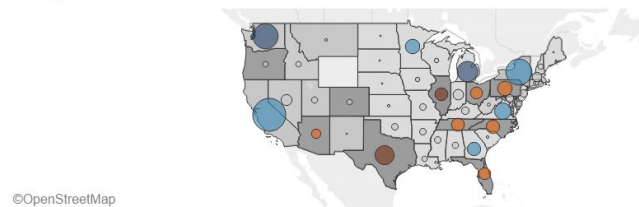
Orange represents loss. Blue represents profit.



Mastering Tableau

Map Loss Leaders

Map



©OpenStreetMap

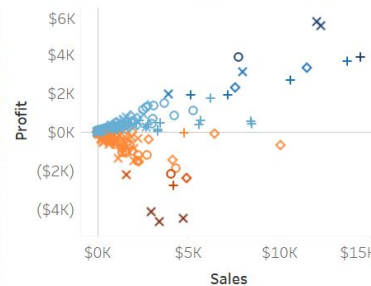
Loss Leader Cities

Orange represents loss. Blue represents profit.

Sta..	City	Avg Dsc..	Discount	Profit	Sales
Texas	Dallas	37.02%	69%	-\$1,984	\$3,507
	Houston	73%		-\$7,520	\$8,148
	San Anto..	64%		-\$4,455	\$4,710
Illinois	Aurora	39.00%	63%	-\$1,475	\$1,983
	Chicago		76%	-\$6,358	\$5,372
Pennsy..	Lancaster	32.86%	70%	-\$106	\$132
	Philadelp..		66%	-\$5,974	\$12,604
Arizona	Phoenix	30.36%	66%	-\$2,779	\$4,183
North ..	Burlington	28.35%	50%	-\$2,159	\$4,045
	Jacksonvi..		67%	-\$474	\$1,177
Florida	Jacksonvi..	29.93%	64%	-\$1,186	\$2,231
Colora..	Aurora	31.65%	67%	-\$576	\$876
Ohio	Lancaster	32.49%	65%	-\$434	\$705
Iowa	Burlington	0.00%	0%	\$3	\$5
Califor..	Lancaster	7.28%	20%	\$41	\$118

Loss Leader Zip

Orange represents loss. Blue represents profit.



Mastering Tableau

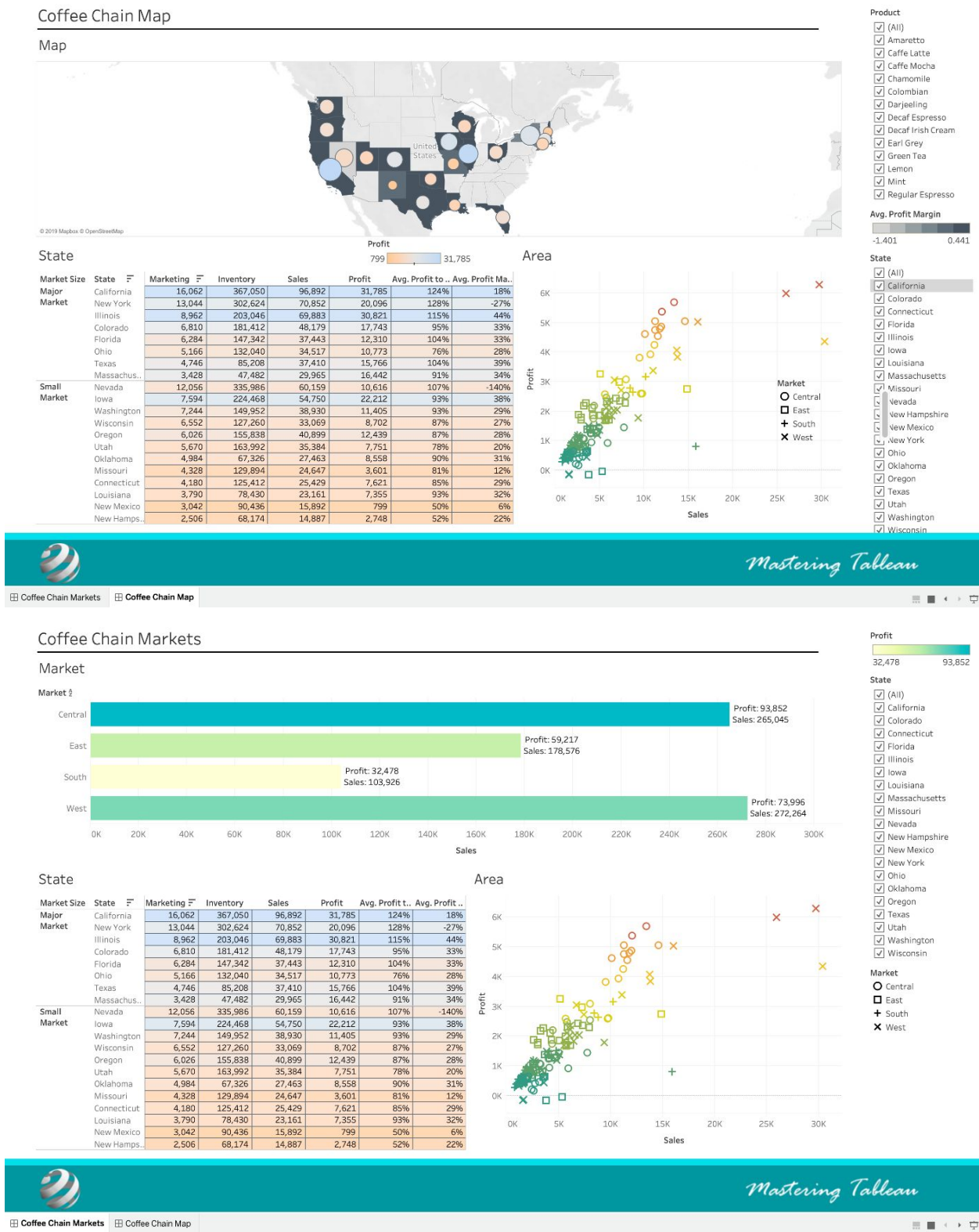
Profit
-17K \$16,922

Sub-Category

- ☐ (All)
- ☐ Accessories
- ☐ Appliances
- ☐ Art
- ☒ Binders
- ☐ Bookcases
- ☐ Chairs
- ☐ Copiers
- ☐ Envelopes
- ☐ Fasteners
- ☐ Furnishings
- ☐ Labels
- ☐ Machines
- ☐ Paper
- ☐ Phones
- ☐ Storage
- ☐ Supplies
- ☒ Tables

Region
x Central
o East
+ South
+ West

Screenshots for application



Coffee Chain Markets

Market

State

Market Size	State	Marketing	Inventory	Sales	Profit	Avg. Profit to ..	Avg. Profit Ma..
Major Market	California	16,062	367,050	96,892	31,785	124%	18%
	New York	13,044	302,624	70,852	20,096	128%	-27%
	Illinois	8,962	203,046	69,883	30,821	115%	44%
	Colorado	6,810	181,412	48,179	17,743	95%	33%
	Florida	6,284	147,342	37,443	12,310	104%	33%
	Ohio	5,166	132,040	34,517	10,773	76%	28%
	Texas	4,746	85,208	37,410	15,766	104%	39%
	Massachus..	3,428	47,482	29,965	16,442	91%	34%
	Nevada	12,056	335,986	60,159	10,616	107%	-140%
	Iowa	7,594	224,468	54,750	22,212	93%	38%
Small Market	Washington	7,244	149,952	38,930	11,405	93%	29%
	Wisconsin	6,552	127,260	33,069	8,702	87%	27%
	Oregon	6,026	155,838	40,899	12,439	87%	28%
	Utah	5,670	163,992	35,384	7,751	78%	20%
	Oklahoma	4,984	67,326	27,463	8,558	90%	31%
	Missouri	4,328	129,894	24,647	3,601	81%	12%
	Connecticut	4,180	125,412	25,429	7,621	85%	29%
	Louisiana	3,790	78,430	23,161	7,355	93%	32%
	New Mexico	3,042	90,436	15,892	799	50%	6%
	New Hamps..	2,506	68,174	14,887	2,748	52%	22%

Area

Profit

State

- ☒ (All)
- ☒ California
- ☒ Colorado
- ☒ Connecticut
- ☒ Florida
- ☒ Illinois
- ☒ Iowa
- ☒ Louisiana
- ☒ Massachusetts
- ☒ Missouri
- ☒ Nevada
- ☒ New Hampshire
- ☒ New Mexico
- ☒ New York
- ☒ Ohio
- ☒ Oklahoma
- ☒ Oregon
- ☒ Texas
- ☒ Utah
- ☒ Washington
- ☒ Wisconsin

Market

- ☒ Central
- ☐ East
- ☐ South
- ☐ West

Mastering Tableau

Coffee Chain Markets

Coffee Chain Map

Description of the new application data

Brief introduction:

The Coffee Chain dataset is from Chapter 8 of Mastering Tableau 2019.2, downloaded from

<https://www.packtpub.com/big-data-and-business-intelligence/mastering-tableau-20191-second-edition>. The data provides operation information(including budget, expenses, sales, profits and so on) of different shops across varying areas of a coffee chain.

Analogous features:

Superstore dataset in the book calculates sales, profit generated by stores in each city of a region, with detailed information like discount and sub-category. In the 2 dashboards created using Superstore data, we can compare sales and profit in city, state and region level.

Just like Superstore data, Coffee Chain data we found also contains operation, product and geographic information of each chain store. Thus, we can make comparison between coffee chain in different areas based on dashboards made from this data.

Insights

For replication

First, among all the stores in the dataset, there's only half of them made a profit. Generally, west region did a better job than other areas, with the highest sales and profit in total.

Second, in this case, profit is not necessarily correlated with sales. There are not a few stores with high amount of sales failed to make a breakeven ultimately.

Finally, stores in categories related to paper(envelopes, labels, paper, copiers) are more likely to generate a profit. In contrast, furniture stores selling tables, chairs and bookcases can be hard to make a breakeven.

For application

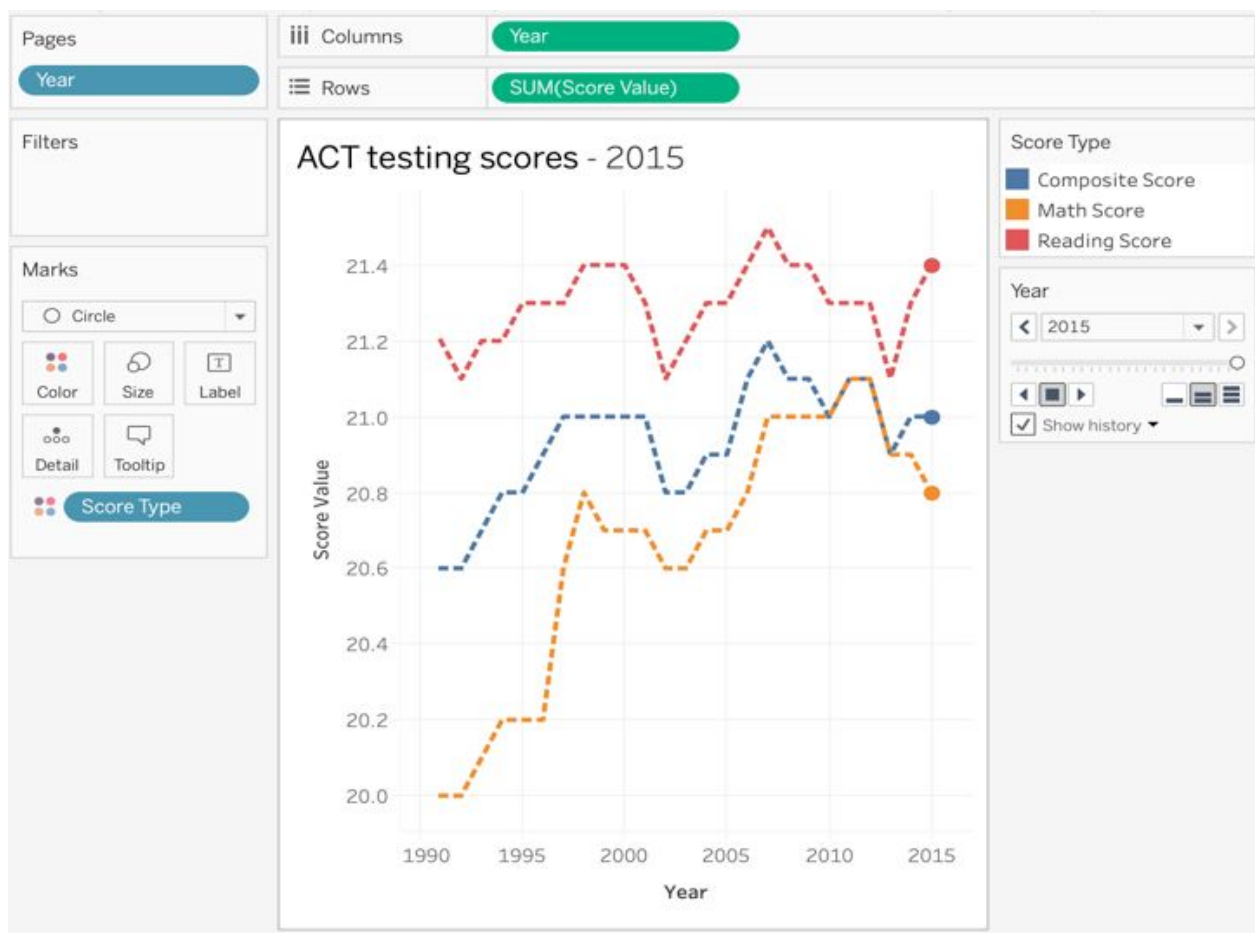
In overall, most coffee stores gain a profit. Central and western area have higher total sales and profit. It might be because the central and western regions have larger marketing size.

In this coffee shops case, profit is highly correlated with sales. Shops selling more tend to make more profit, no matter what types of coffee they sell.

Another finding is that among all the products sold by the coffee chain, mint and regular espresso are least popular. Regular espresso is only available in east area..

D. Chapter 10, page 322, creating an animation with Tableau

Screenshots for replication



Screenshots for application

(Suicide rate/100k pop in United States 1985-2015)






Description of the new application data

Brief introduction:

The dataset is downloaded from

<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016>, which

includes 4 columns: country (select “United States”, year, age group, suicide rate/100k pop.

	country	# year	A age	# suicides/100k pop
			<div>15-24 years 17%</div> <div>35-54 years 17%</div> <div>Other (4) 67%</div>	
1	United States	1985	35-54 years	22.77
2	United States	1985	35-54 years	7.58
3	United States	1986	35-54 years	23.57
4	United States	1986	35-54 years	8.06
5	United States	1987	35-54 years	23.28
6	United States	1987	35-54 years	7.73
7	United States	1988	35-54 years	22.4
8	United States	1988	35-54 years	7.3
9	United States	1989	35-54 years	22.61
10	United States	1989	35-54 years	6.91

Analogous feature:

The ACT test score data in the book includes 3 columns, which are score type (group), score value (value) and year (time). With the data we can create an animation and see the different types of score value change along with time.

Similarly, the new dataset I chose also includes 3 columns, which are age group (group), suicide rate/100k pop (value), year (time). With the data we can create an animation and see the different suicide rate in different age group change along with time.

Insight

For replication

We saw all kinds of ACT testing scores increase from 1990 to 2015 although fluctuations exist, among which math score has the maximum increase but still ranks third by the end of 2015 after reading score and composite score. That is to say, reading is easier for students compared to math and composing.

For replication

From 1985 to 2015, the suicide rate of United States decreases and all of them are below 45 (/100k pop), which is a good sign. Generally speaking, the suicide rate correlated with age, the older the higher the suicide rate is, except for the age group of 35-54, which is higher than the age group of 55-74 after 1995. Maybe we should pay special attention to the 35-54 age group when promoting suicide prevention measures.

The animation via tableau enables us to gain a clearer and more direct view of the dynamic trend throughout the years and see the components of interested variables in details. However, an animation can certainly make for engaging communication but does not work well for deeper dives into the data. For example, we cannot figure out the reason behind why people among 35-54 have a relatively high suicide rate. Therefore, other visualization and analysis tools should be considered to dig further.