US Monthly Retail Trade

Data Analysis and Visualization by Rshiny Web App

Outline

1. Introduction
2. Data Set
3. Methods
4. Reference
5. Run App & Discussion

1. Introduction/Background

How dose Census Collect and Store the Data？

The Advance Monthly and Monthly Retail Trade Surveys (MARTS and MRTS), the Annual Retail Trade Survey (ARTS), and the Quarterly E-Commerce Report work together to produce the most comprehensive data available on retail economic activity in the United States.

Regular quality control and verification takes place between MARTS, MRTS, and ARTS annually and between these programs and the Economic Census of Retail Trade every five years. Each year when annual data become available, census compare and resolve differences between the data collected on the monthly and annual surveys. It refer to this process as the monthly-to-annual reconciliation. At the same time, census benchmark the monthly estimate using results of the annual survey. ARTS estimates are then benchmarked to data maintained by the Economic Census of Retail Trade. This process of benchmarking retail data over all four programs ensures consistency in its estimates.

Thank Census for doing this regulary:

1. Collect and store data.
2. Provide updated data.
3. Adjust data for seasonal, holiday, and trading-day differences, but not for price changes.

However?

What is the Problem & What I contribute

1. I think people want to see insights based on US Monthly Retail Trade data. It is like reading a history book digging this wide year range data. Especially retail economy is close to life. We might have a further understanding of something beyond math, CS, or even economy.
2. Doing some details is more fun than glancing the whole data table. For example, users will learn how many business in retail industry, they can also compare different kinds of retail business, and forecast trends in the future.
3. Users interested in the topic a quick guide by running App and selecting preferences.

2. Data Set

Data Set

1. Background

   Sales data ranges from Jan1992 to Jan2023, including various retail business kinds.

2. Structure

   It is an xlsx file with 31 sheets, each sheet contains a data report.

   Processing data reasonably is really important.

A lot of work to do!

441	Motor vehicle and parts dealers
4411,4412	Automobile and other motor vehicle dealers
4413	Automotive parts, acc., and tire stores
442,443	Furniture, home furn, electronics, and appliance stores
442	Furniture and home furnishings stores
443	Electronics and appliance stores
444	Building mat. and garden equip. and supplies dealers
4441	Building mat. and supplies dealers
445	Food and beverage stores
4451	Grocery stores
4453	Beer, wine and liquor stores
446	Health and personal care stores
44611	Pharmacies and drug stores
447	Gasoline stations
448	Clothing and clothing access. stores
4481	Clothing stores
44811	Men's clothing stores
44812	Women's clothing stores
4482	Shoe stores
44831	Jewelry stores
451	Sporting goods, hobby, musical instrument, and book stores
452	General merchandise stores
4521	Department stores
4529	Other general merchandise stores
45291	Warehouse clubs and superstores
45299	All other gen. merchandise stores
453	Miscellaneous stores retailers
454	Nonstore retailers
4541	Electronic shopping and mail order houses
45431	Fuel dealers
722	Food services and drinking places

See, there are too many similar kinds of business, which is hard to analysis.

Here need to read and process multiple data reports from sheets. Data type, data structure should be suitable for analysis, plot and visualization. Also need to convert data structure to ts for forecast.

Estimates of Monthly Retail and Food Services Sales by Kind of Business: 2006

[Estimates are shown in millions of dollars and are based on data from the Monthly Retail Trade Survey, Annual Retail Trade Survey, Service Annual Surve

NAICS Code	Kind of Business	Jan. 2006	Feb. 2006	Mar. 2006	Apr. 2006	May
	NOT ADJUSTED					
	Retail and food services sales, total	318,546	314,051	361,993	351,667	
	Retail sales and food services excl motor vehicle and parts	252,809	246,599	279,239	275,487	
	Retail sales and food services excl gasoline stations	287,746	284,798	328,106	315,665	
	Retail sales and food services excl motor vehicle and parts and gasolin	222,009	217,346	245,352	239,485	
	Retail sales, total	286,152	282,417	326,153	316,526	
	Retail sales, total (excl. motor vehicle and parts dealers)	220,415	214,965	243,399	240,346	
	GAFO(1)	78,420	79,008	86,862	85,932	
441	Motor vehicle and parts dealers	65,737	67,452	82,754	76,180	
4411,4412	Automobile and other motor vehicle dealers	60,247	62,018	76,416	70,287	
4411	Automobile dealers	56,440	57,559	69,930	63,333	
44111	New car dealers	50,629	51,314	62,921	56,863	
44112	Used car dealers	5,811	6,245	7,009	6,470	
4413	Automotive parts, acc., and tire stores	5,490	5,434	6,338	5,893	
442,443	Furniture, home furn, electronics, and appliance stores	17,370	16,693	18,092	16,346	
442	Furniture and home furnishings stores	8,595	8,457	9,412	8,556	
4421	Furniture stores	4,794	4,804	5,154	4,647	
4422	Home furnishings stores	3,801	3,653	4,258	3,909	
44221	Floor covering stores	1,664	1,712	1,985	1,851	
442299	All other home furnishings stores	2,008	1,810	2,112	1,908	
443	Electronics and appliance stores	8,775	8,236	8,680	7,790	
443141	Household appliance stores	1,325	1,208	1,457	1,263	

2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996

3. Methods

Methods

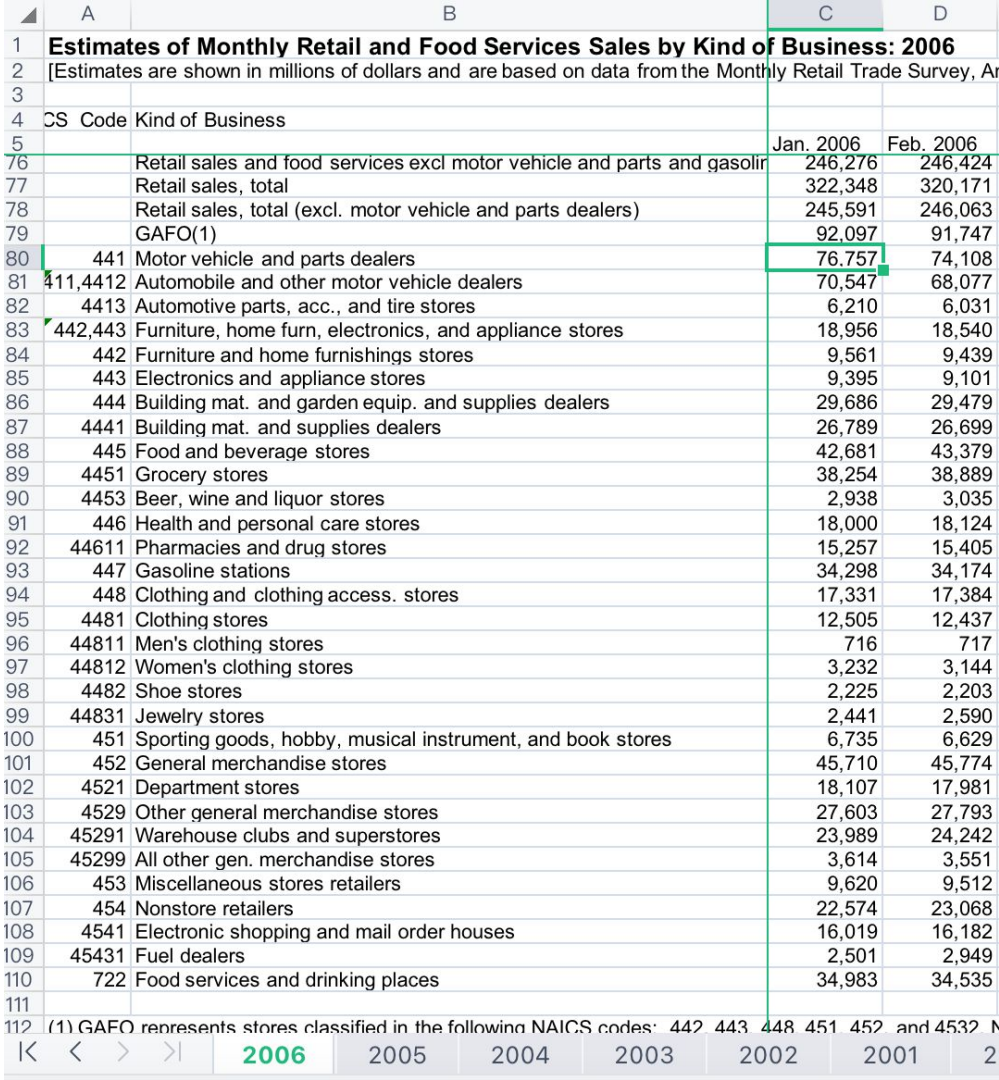
1. Import Data
2. Tidy Data/Data Transformation/Data Wrangle
   Relational Data(not involved)
3. Algorithm/function
4. Interactive Widgets/Frond End
5. Data Visualization
6. Other Novel Functions, hover interaction, etc. (optional)

Methods...

1. It is not only a platform for visualizing data, but more importantly, it implemented the interactive function of user self-analysis. Users can select and compare unique business kinds based on their own insights in the line plot and do analysis combined with the other three plots.
2. Users can have fun exploring forecast implementation, which provides a tiny but complete forecast panel for all businesses kinds.
3. Keeping overall data distribution stationery is useful when digging into details in other plots.

Methods 1.Import Data

- Download .xlsx file from url
- Find data structures and certain patterns
- lapplyfunction/excel_sheets /read_excel are my good friends!



	A	B	C	D
1	**Estimates of Monthly Retail and Food Services Sales by Kind of Business: 2006**			
2	[Estimates are shown in millions of dollars and are based on data from the Monthly Retail Trade Survey, Ar			
3				
4	CS Code	Kind of Business		
5			Jan. 2006	Feb. 2006
76		Retail sales and food services excl motor vehicle and parts and gasolir	246,276	246,424
77		Retail sales, total	322,348	320,171
78		Retail sales, total (excl. motor vehicle and parts dealers)	245,591	246,063
79		GAFO(1)	92,097	91,747
80	441	Motor vehicle and parts dealers	76,757	74,108
81	411,4412	Automobile and other motor vehicle dealers	70,547	68,077
82	4413	Automotive parts, acc., and tire stores	6,210	6,031
83	442,443	Furniture, home furn, electronics, and appliance stores	18,956	18,540
84	442	Furniture and home furnishings stores	9,561	9,439
85	443	Electronics and appliance stores	9,395	9,101
86	444	Building mat. and garden equip. and supplies dealers	29,686	29,479
87	4441	Building mat. and supplies dealers	26,789	26,699
88	445	Food and beverage stores	42,681	43,379
89	4451	Grocery stores	38,254	38,889
90	4453	Beer, wine and liquor stores	2,938	3,035
91	446	Health and personal care stores	18,000	18,124
92	44611	Pharmacies and drug stores	15,257	15,405
93	447	Gasoline stations	34,298	34,174
94	448	Clothing and clothing access. stores	17,331	17,384
95	4481	Clothing stores	12,505	12,437
96	44811	Men's clothing stores	716	717
97	44812	Women's clothing stores	3,232	3,144
98	4482	Shoe stores	2,225	2,203
99	44831	Jewelry stores	2,441	2,590
100	451	Sporting goods, hobby, musical instrument, and book stores	6,735	6,629
101	452	General merchandise stores	45,710	45,774
102	4521	Department stores	18,107	17,981
103	4529	Other general merchandise stores	27,603	27,793
104	45291	Warehouse clubs and superstores	23,989	24,242
105	45299	All other gen. merchandise stores	3,614	3,551
106	453	Miscellaneous stores retailers	9,620	9,512
107	454	Nonstore retailers	22,574	23,068
108	4541	Electronic shopping and mail order houses	16,019	16,182
109	45431	Fuel dealers	2,501	2,949
110	722	Food services and drinking places	34,983	34,535
111				
112	(1) GAFO represents stores classified in the following NAICS codes: 442, 443, 448, 451, 452, and 4532, N			

2006 2005 2004 2003 2002 2001

Methods 2. tidy/transformation/wran

- rename column names
- convert data type(date/yearmon/factor)
- pivot table(all groups/one group ts)
- group large kinds based on definition
- transform value to log
- paste a lot but help a lot as well
- build for loop/if else functions help a lot
- create several data subset is necessary.

Jan. 2006	Feb. 2006	Mar. 2006	Apr. 2006
318,546	314,051	361,993	351,667
252,809	246,599	279,239	275,487
287,746	284,798	328,106	315,665
222,009	217,346	245,352	239,485
286,152	282,417	326,153	316,526
220,415	214,965	243,399	240,346

441	Motor vehicle and parts dealers
4411,4412	Automobile and other motor vehicle dealers
4413	Automotive parts, acc., and tire stores
442,443	Furniture, home furn, electronics, and applian
442	Furniture and home furnishings stores
443	Electronics and appliance stores
444	Building mat. and garden equip. and supplies
4441	Building mat. and supplies dealers
445	Food and beverage stores
4451	Grocery stores
4453	Beer, wine and liquor stores
446	Health and personal care stores
44611	Pharmacies and drug stores
447	Gasoline stations
448	Clothing and clothing access. stores
4481	Clothing stores
44811	Men's clothing stores
44812	Women's clothing stores
4482	Shoe stores
44831	Jewelry stores
451	Sporting goods, hobby, musical instrument, ar
452	General merchandise stores
4521	Department stores
4529	Other general merchandise stores
45291	Warehouse clubs and superstores
45299	All other gen. merchandise stores
453	Miscellaneous stores retailers
454	Nonstore retailers
4541	Electronic shopping and mail order houses
45431	Fuel dealers
722	Food services and drinking places

Methods 2. Data Transformation(An example)

This is the dataframe from the very begaining after import

Name	Type	Value
df_nd	list [31]	List of length 31
2022	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r
2021	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r
2020	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r
2019	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r
2018	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r
2017	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r
2016	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r
2015	list [31 x 12] (S3: tbl_df, tbl, c	A tibble with 31 r

Methods 2. Data Transformation(An example)

After process: bind_cols/remove nan cols/rows/aggregate/melt/datatype…

group	month_year	sales	month	year	month_num	dt	dt_month_year
art	Jan.1992	3385	Jan	1992	1	1992-01-01	Jan 1992
clothing	Jan.1992	20713	Jan	1992	1	1992-01-01	Jan 1992
eshop	Jan.1992	9183	Jan	1992	1	1992-01-01	Jan 1992
fuel	Jan.1992	14085	Jan	1992	1	1992-01-01	Jan 1992
furniture	Jan.1992	36160	Jan	1992	1	1992-01-01	Jan 1992
general	Jan.1992	49821	Jan	1992	1	1992-01-01	Jan 1992
grocery	Jan.1992	59985	Jan	1992	1	1992-01-01	Jan 1992
health	Jan.1992	13882	Jan	1992	1	1992-01-01	Jan 1992
restaurant	Jan.1992	17170	Jan	1992	1	1992-01-01	Jan 1992
vehicle	Jan.1992	66956	Jan	1992	1	1992-01-01	Jan 1992
art	Feb.1992	3516	Feb	1992	2	1992-02-01	Feb 1992
clothing	Feb.1992	20689	Feb	1992	2	1992-02-01	Feb 1992
eshop	Feb.1992	8742	Feb	1992	2	1992-02-01	Feb 1992
fuel	Feb.1992	13778	Feb	1992	2	1992-02-01	Feb 1992
furniture	Feb.1992	36480	Feb	1992	2	1992-02-01	Feb 1992
general	Feb.1992	50442	Feb	1992	2	1992-02-01	Feb 1992

Methods 2. Data Transformation(An example)

Prepare ts dataset for forecast: store all groups in list

Name	Type	Value
df_list	list [10]	List of length 10
df_art	double [373]	3385 3516 3482 3535 3512 3549 ...
df_clothing	double [373]	20713 20689 20520 21023 20827 21096 ...
df_eshop	double [373]	9183 8742 8495 9084 9424 9718 ...
df_fuel	double [373]	14085 13778 13966 14174 14361 14307 ...
df_furniture	double [373]	36160 36480 36616 36386 36445 36696 ...
df_general	double [373]	49821 50442 50253 50636 50757 50654 ...
df_grocery	double [373]	59985 59865 60127 60253 60488 60554 ...
df_health	double [373]	13882 13883 13902 14050 13969 13875 ...
df_restaurant	double [373]	17170 16990 16916 16677 16602 16277 ...
df_vehicle	double [373]	66956 67634 66672 67552 68714 69562 ...

```
> df_list[["df_art"]]
  [1] 3385 3516 3482 3535 3512 3549 3586 3512 3566 3607 3626 3800 3648 3585 3608 3749 3746 3702 3810
 [20] 3804 3825 3817 3930 4018 3981 4095 4131 4035 4119 4135 4111 4249 4306 4347 4221 4255 4436 4298
 [39] 4347 4345 4380 4421 4429 4462 4441 4414 4532 4505 4600 4485 4548 4628 4432 4609 4655 4738 4690
 [58] 4769 4692 4940 4748 4666 4698 4725 4763 4838 4786 4745 4827 4878 5015 4915 4834 5027 5008 5036
 [77] 4986 5083 5032 5057 5047 5108 5093 5154 5122 5292 5273 5277 5328 5367 5326 5290 5385 5373 5414
 [96] 5455 5232 5541 5635 5649 5757 5686 5800 5633 5728 5626 5615 5601 5386 5715 5723 5658 5679 5621
[115] 5703 5775 5672 5798 6066 5797 5775 5711 5717 5799 5731 5670 5666 5689 5852 5839 5985 5809 5794
[134] 5576 5660 5709 5748 5908 5808 6132 5954 5841 6011 6013 6087 6093 6067 5941 5985 6017 6124 6192
[153] 6111 6094 6140 6122 6192 6269 6137 6318 6232 6135 6239 6274 6155 6377 6287 6462 6735 6629 6610
[172] 6580 6503 6524 6461 6412 6690 6589 6461 6566 6489 6494 6779 6640 6805 6850 6936 6780 6749 6834
[191] 6940 6756 6818 6751 6733 6809 6949 6879 6858 6931 6536 6426 6431 6470 6661 6535 6360 6514 6453
[210] 6473 6455 6628 6484 6408 6307 6523 6507 6412 6515 6409 6496 6497 6516 6441 6469 6579 6723 6588
[229] 6565 6611 6566 6608 6542 6585 6462 6661 6686 6605 6544 6353 6575 6683 6815 6859 6869 6797 6813
[248] 6715 6720 6749 6675 6828 7012 7021 6899 6795 6862 6913 6938 6713 6877 6860 6875 6861 6488 6795
[267] 6803 6894 6860 6950 6922 7122 7043 7156 7147 7066 6939 6977 7061 7084 7121 7069 7124 7158 7188
```

Methods 2. Data Transformation(An example)

Prepare ts dataset for forecast: ts loop all groups

Name	Type	Value
ts_list	list [10]	List of length 10
ts_df_art	double [373] (S3: ts)	3385 3516 3482 3535 3512 3549 ...
ts_df_clothing	double [373] (S3: ts)	20713 20689 20520 21023 20827 21096 ...
ts_df_eshop	double [373] (S3: ts)	9183 8742 8495 9084 9424 9718 ...
ts_df_fuel	double [373] (S3: ts)	14085 13778 13966 14174 14361 14307 ...
ts_df_furniture	double [373] (S3: ts)	36160 36480 36616 36386 36445 36696 ...
ts_df_general	double [373] (S3: ts)	49821 50442 50253 50636 50757 50654 ...
ts_df_grocery	double [373] (S3: ts)	59985 59865 60127 60253 60488 60554 ...
ts_df_health	double [373] (S3: ts)	13882 13883 13902 14050 13969 13875 ...
ts_df_restaurant	double [373] (S3: ts)	17170 16990 16916 16677 16602 16277 ...
ts_df_vehicle	double [373] (S3: ts)	66956 67634 66672 67552 68714 69562 ...

```
> View(ts_list)
> ts_list[["ts_df_art"]]
      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
1992 3385 3516 3482 3535 3512 3549 3586 3512 3566 3607 3626 3800
1993 3648 3585 3608 3749 3746 3702 3810 3804 3825 3817 3930 4018
1994 3981 4095 4131 4035 4119 4135 4111 4249 4306 4347 4221 4255
1995 4436 4298 4347 4345 4380 4421 4429 4462 4441 4414 4532 4505
1996 4600 4485 4548 4628 4432 4609 4655 4738 4690 4769 4692 4940
1997 4748 4666 4698 4725 4763 4838 4786 4745 4827 4878 5015 4915
1998 4834 5027 5008 5036 4986 5083 5032 5057 5047 5108 5093 5154
1999 5122 5292 5273 5277 5328 5367 5326 5290 5385 5373 5414 5455
2000 5232 5541 5635 5649 5757 5686 5800 5633 5728 5626 5615 5601
2001 5386 5715 5723 5658 5679 5621 5703 5775 5672 5798 6066 5797
2002 5775 5711 5717 5799 5731 5670 5666 5689 5852 5839 5985 5809
2003 5794 5576 5660 5709 5748 5908 5808 6132 5954 5841 6011 6013
2004 6087 6093 6067 5941 5985 6017 6124 6192 6111 6094 6140 6122
2005 6192 6269 6137 6318 6232 6135 6239 6274 6155 6377 6287 6462
2006 6735 6629 6610 6580 6503 6524 6461 6412 6690 6589 6461 6566
2007 6489 6494 6779 6640 6805 6850 6936 6780 6749 6834 6940 6756
2008 6818 6751 6733 6809 6949 6879 6858 6931 6536 6426 6431 6470
2009 6661 6535 6360 6514 6453 6473 6455 6628 6484 6408 6307 6523
```

Methods 3. Functions

- Built 3 functions to process data for different use and sitution.
- Built 2 functions to control interactive plot.
- Built 3 functions to control interactive forecast and plot.

- Built time series models for 10 kinds of business predicting 24 months, which perform well and reach small errors.

  Holt-Winters Exponential Smoothing Model was selected because this time series can be described using an additive model with increasing trend and seasonality, Holt-Winters is a good and simple choice to make short-term forecasts.

Holt-Winters

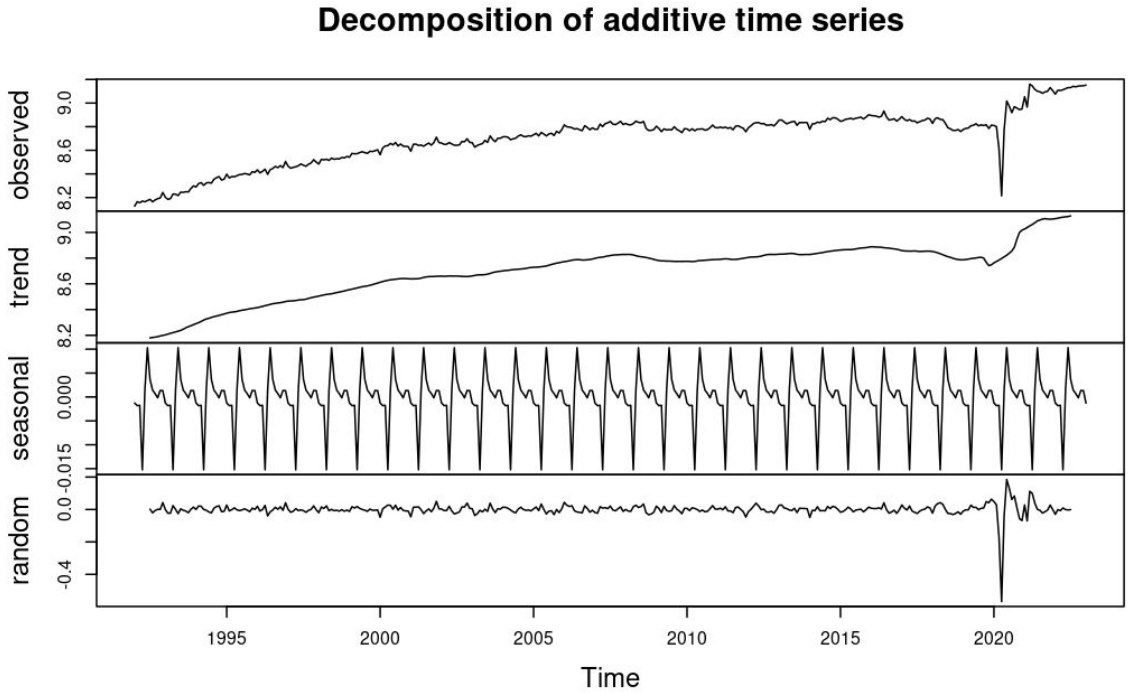
Holt-Winters exponential smoothing estimates the level, slope and seasonal component at the current time point.

Smoothing is controlled by three parameters: alpha, beta, and gamma, for the estimates of the level, slope b of the trend component, and the seasonal component, respectively, at the current time point. The parameters alpha, beta and gamma all have values between 0 and 1, and values that are close to 0 mean that relatively little weight is placed on the most recent observations when making forecasts of future values.

Decomposition interestingly separates out the 3 main components that make up the time series:

1. trend: the long-term trends in the data
2. seasonal: the repeated seasonal signal adder
3. random: the "left-over" components that aren't expected from the seasonality or trend components.

Decompose



Methods 4. Interactive Widgets/Frond End

Built several interactive widgets for users to select.

For example:

- “year_range” to control plotting time range,
- “grList1” to select business kinds for plotting,
- “compareList” to select two business kinds for comparing their corr,
- “forecastbusiness” to select one retail kind for forecast.

Methods 5. Data Visualization

- Plot 2 overview distribution figures.
- Plot 2 zoom in figures to see details.
- Plot 4 time series figures.

4. Reference

Reference

1. Census: https://www.census.gov/retail/about_the_surveys.html
2. Rshiny: https://shiny.rstudio.com/
3. Github:https://github.com/rstudio/shinydashboard
4. Slides and homeworks in IE6600 class!

5. Run App & Discussion

Run App

Run App: https://hiaudrey.shinyapps.io/shinyapp/

Data: https://www.census.gov/retail/mrts/www/mrtssales92-present.xlsx

Github: https://github.com/HiAudery/Cider/