Visualizing London House Prices in Tableau

1. Introduction

The main purpose of this exercise is to give you an opportunity to analyse a large data set and present what you find through a data story. The main question to explore is the impact of Brexit and Covid on London House Prices using visualization tools. You will use Tableau software to facilitate the analysis. You will also answer some additional related questions posed in the case.

Brexit: Brexit is the withdrawal of the United Kingdom (UK) from the European Union (EU). Following a June 2016 referendum, in which 51.9% voted to leave, the UK government formally announced the country's withdrawal in March 2017, starting a two-year process that was due to conclude with the UK withdrawing on 29 March 2019. As the UK parliament voted against or failed to ratify the negotiated withdrawal agreements, that deadline has been extended three times, and UK left EU officially on 31/12/2020.

Throughout this workshop you can find example visualizations and screencasts that demonstrate how to use different tools in Tableau. Some of these are prepared using different versions of the data set you have access to. Therefore, you may not get identical results to those you see in these resources after you implement the same steps.

Your assignment is based on this workshop so please make sure to save your work frequently.

2. Data sources

Publicly available (https://www.gov.uk/government/collections/price-paid-data)
HM Land Registry's Price Paid Data tracks the property sales in England and Wales that are submitted to them for registration. The sales data is available between 1998-Present. See Section 3 below for the data dictionary and details of the types of properties included in the data set. We only consider the data for London.

We merge this data set with (publicly available) Energy Performance Certificate (EPC) data. This database contains more information about each property including size, number of bedrooms, and the energy ratings. However, this data set only runs from 2008 to 2020 and does not contain all the properties in the HM Land Registry's database. Finally, we merged the resulting data set with data about each post code.

Data Cleaning: Data is already cleaned to the best of my ability. If you see any potential problems, feel free to filter out the problematic data. You may also need to filter out

property_type ***Other***, given as an ***O*** in the dataset, that includes non-residential properties like the LBS building!

Download data from

https://www.dropbox.com/scl/fi/r52pw3xsxmajydreluowz/london-pp-epc.csv.zip?rlkey=o3m3df51wltq98dfvk1okbxdh&dl=1

to get a zipped (*.zip) file with London house prices between January, 2000 and August, 2023. Please note that the data from the last few months of 2021 may not be complete because the website is not updated real time. Please extract/unzip the file so you have a CSV (Comma Separated Values) file that you can import into Tableau.

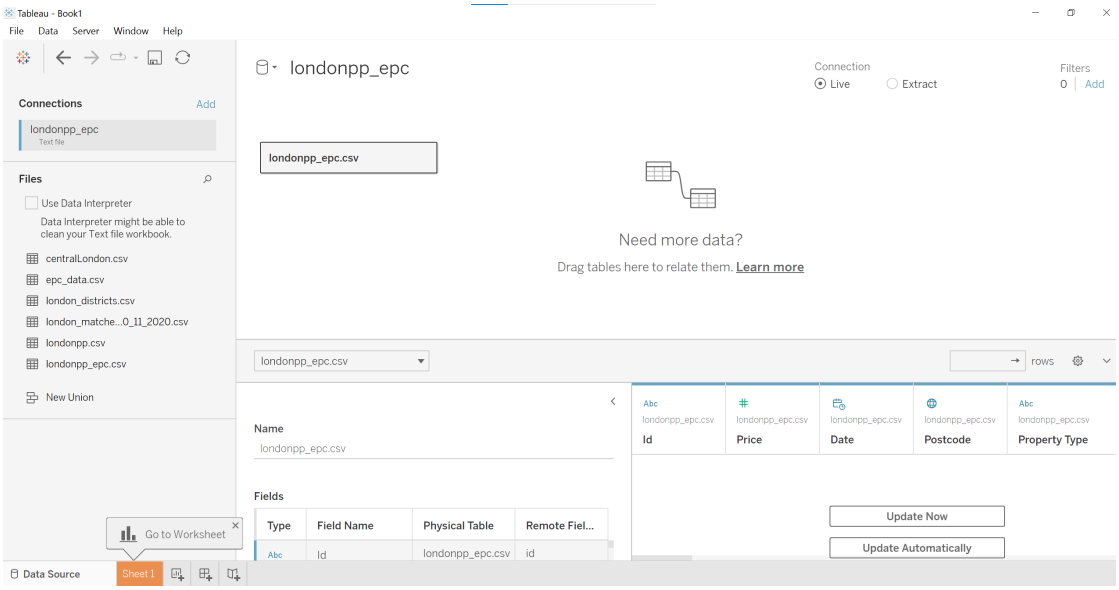
3. Data Dictionary

Price Paid Data	
ID	A reference number which is generated automatically recording each published sale. The number is unique and will change each time a sale is recorded.
Price	Sale price stated on the transfer deed.
Date	Date when the sale was completed, as stated on the transfer deed.
Postcode	This is the postcode used at the time of the original transaction.
postcode_short	First part of postcode
Property Type	D = Detached, S = Semi-Detached, T = Terraced, F = Flats/Maisonettes, O = Other Note that: - we only record the above categories to describe property type, we do not separately identify bungalows. - end-of-terrace properties are included in the Terraced category above. - 'Other' is only valid where the transaction relates to a property type that is not covered by existing values.
Whether_old_new	Indicates the age of the property and applies to all price paid transactions, residential and non-residential. Y = a newly built property, N = an established residential building
freehold_or_leasehold	Relates to the tenure: F = Freehold, L= Leasehold etc. Note that HM Land Registry does not record leases of 7 years or less in the Price Paid Dataset.
Adress 1	
Adress 2	
Adress 3	
town	
local_aut	
county	
district	
pricePerSqmtr	Price per square meter
Energy Performance of Buildings Data	
total_floor_area	The total useful floor area is the total of all enclosed spaces measured to the internal face of the external walls
number_habitable_rooms	Habitable rooms include any living room, sitting room, dining room, bedroom, study and similar; and also a non-separated conservatory
current_energy_rating	Current energy rating converted into a linear 'A to G' rating (where A is the most energy efficient and G is the least energy efficient)
co2_emissions_current	CO_2 emissions per year in tonnes/year
co2_emissions_potential	Estimated value in Tonnes per Year of the total CO_2 emissions produced by the Property in 12 month period.
energy_consumption_current	Current estimated total energy consumption for the property in a 12 month period (kWh/m2). Displayed on EPC as the current primary energy use per square metre of floor area
energy_consumption_potential	Estimated potential total energy consumption for the Property in a 12 month period. Value is Kilowatt Hours per Square Metre (kWh/m²)
windows_energy_eff	Energy efficiency rating. One of: very good; good; average; poor; very poor. On actual energy certificate shown as one to five star rating.
tenure	Describes the tenure type of the property. One of: Owner-occupied; Rented (social); Rented (private).
Postcode data	
population	Population of the area covered by the Postcode (from the 2011 census)
households	Number of Households in the area covered by the Postcode (from the 2011 census)
latitude	Latitude of centroid of the Postcode for this row in decimal format i.e. 51.50205
longitude	Longditude of centroid of the Postcode for this row in decimal format. i.e -0.07864. Negative values are those to the West of the zero (Greenwich) meridian
altitude	Height above sea level measured in Metres
london_zone	Transport for London (TfL) Travel Zone indicator (London area only)
user_type	Shows whether the postcode is a small or large user. 0 = small user (addresses receive less than 25 items of mail per day) 1 = large user (addresses receive more than 25 items of mail per day)
nearest_station	The nearest train station to the postcode. For London, also includes Underground and tram stops
distance_to_station	The distance in kilometres to the nearest station from the postcode
water_company	The name of the water company responsible for this postcode
average_income	Average household income of the MSOA that the postcode is located in

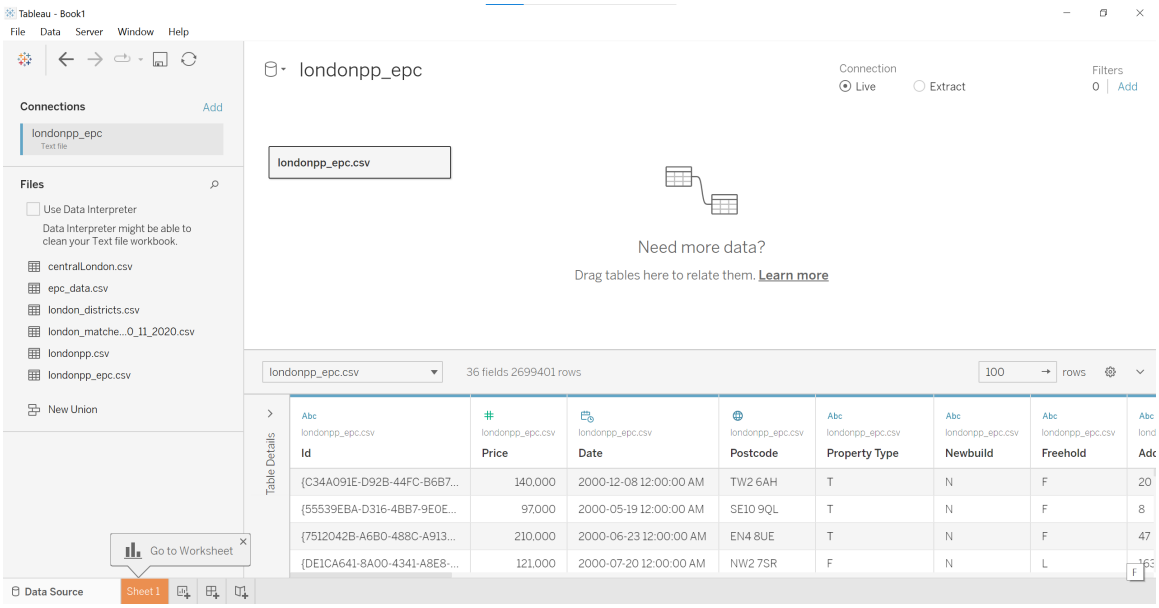
Figure 1 Data Dictionary

4. Connect to data

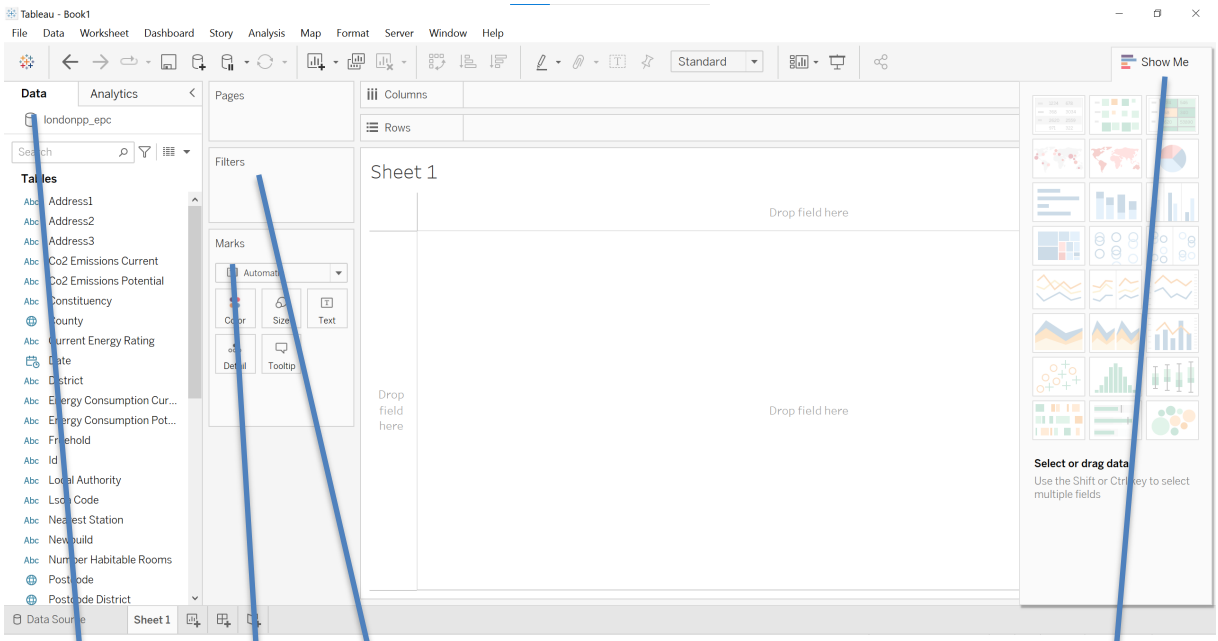
Once you open Tableau, you can load the data files on the landing page. Tableau allows the input data to be in different file formats and Price Paid Data is a comma separated values (csv) file. Load the data by going to Tableau -> ***Connect… -> To a file -> Text File***… Once uploading is complete you should see the preview of the data, similar to the screen capture below.



Click ***Update Now*** if you cannot see the data—this may take a while to load, as this is a fairly large file, but once it does you can preview the data as it were an Excel file.



Once data upload is complete click on Sheet 1 in the bottom. You will see a page similar to the screen shot below. The instructions below use the terminology outlined in this figure.



Marks

Visualisation options

Data Summary

Filters

Further details can be found at https://help.tableau.com/current/pro/desktop/en-us/environment_workspace.htm

5. Get to know your data

We begin our analysis with exploring the price data.

Investigate if the price data has any outliers. What is the range of prices for London Houses? What is the distribution of the prices? Can you find the most expensive property in London and see what it looks like in Google maps?

First take a look at the variables that appear in the bottom half of the Data Source Pane. Familiarize yourself with the data but note that we will not use all the variables in this data set in this exercise.

We used R to quickly manipulate the data and have a look at the 25 most expensive properties, listed below….

	id	price	date	postcode	property_type	newbuild	freehold
1	{C6209F60-4924-295E-E053-6C04A8C0DDCC}	630000000	2021-03-29	CR7 6AZ	O	N	F
2	{582D0637-EE28-8F22-E053-6C04A8C01BAC}	594300000	2017-07-31	W1U 8EW	O	N	L
3	{666758D7-A512-3363-E053-6B04A8C0D74E}	569200000	2018-02-08	W1J 7BT	O	N	F
4	{7E86B6FB-BF88-458C-E053-6B04A8C0C84C}	448500000	2018-09-21	SE1 9JZ	O	N	F
5	{CD5A9DCC-21D7-310A-E053-6C04A8C00A1F}	422500000	2021-05-20	UB10 0SD	T	N	F
6	{8CAC1318-F392-0253-E053-6B04A8C08E51}	411500000	2019-04-04	EC2V 7PG	O	N	L
7	{55BDCAE6-DC9C-521D-E053-6B04A8C0DD7A}	370000000	2016-10-31	SW1H 0BG	O	N	F
8	{B82222ED-A781-6691-E053-6B04A8C02FB2}	366180000	2020-10-08	W1S 2UB	O	N	F
9	{B82222ED-8D16-6691-E053-6B04A8C02FB2}	366180000	2020-10-08	W1S 2UD	O	N	L
10	{98C75472-BCE7-72E9-E053-6B04A8C042F0}	337000000	2019-04-04	EC4V 5AE	O	N	F
11	{773788C3-9644-2CE4-E053-6C04A8C05E57}	325169089	2018-06-29	SW1A 2EU	O	N	L
12	{965B6D92-4C16-95E4-E053-6C04A8C07729}	315000000	2019-03-27	SW6 2YA	O	N	F
13	{B5755FFE-98AA-7D01-E053-6C04A8C036B6}	303470000	2020-06-30	EC1V 3PS	O	N	L
14	{93E6821F-1B9B-40FD-E053-6B04A8C0C1DF}	302401851	2019-03-15	EC4V 5AJ	O	N	L
15	{7E86B6FB-B551-458C-E053-6B04A8C0C84C}	302000000	2018-12-07	WC1B 4JB	O	N	F
16	{BC8936BC-406A-0E2C-E053-6C04A8C0DBF4}	296280000	2020-12-16	E16 2QQ	O	N	F
17	{BA558B33-80A0-76EF-E053-6B04A8C0B4B7}	286000000	2020-09-24	WC1N 3PA	O	N	F
18	{55BDCAE6-C8A8-521D-E053-6B04A8C0DD7A}	280000000	2017-04-19	W14 8UX	O	N	F
19	{919FEC06-343C-9A90-E053-6C04A8C0A300}	276570000	2019-06-11	E20 1DB	O	N	L
20	{5F54B81E-2C91-2B45-E053-6B04A8C01FB0}	270798594	2017-10-12	EC3M 6DE	O	N	F
21	{64342BFF-5BCE-422C-E053-6C04A8C0FB8A}	263898965	2017-09-26	SW1W 8RF	O	N	F
22	{55BDCAE6-C6A6-521D-E053-6B04A8C0DD7A}	263100000	2016-11-24	W1D 1BY	O	N	F
23	{965B6D92-41B4-95E4-E053-6C04A8C07729}	262890000	2019-09-27	E14 5HU	O	N	L
24	{85866A65-9A1C-143F-E053-6B04A8C06A15}	261710920	2018-12-21	W1A 1AE	O	N	L
25	{79A74E22-7C45-1289-E053-6B04A8C01627}	261500000	2018-09-25	W11 4UL	O	N	F

… as well as the 25 least expensive properties, again shown below.

	id	price	date	postcode	property_type	newbuild	freehold
1	{2A289E9D-A56A-CDC8-E050-A8C063054829}	1	2000-11-23	N16 7PT	F	N	L
2	{2A289E9C-DFEE-CDC8-E050-A8C063054829}	1	2002-08-01	CR0 6DQ	F	N	L
3	{2FD36065-20A0-4BF8-E050-A8C0620562B1}	1	2016-02-17	HA0 3NG	T	N	L
4	{B32EBB13-3C67-3C99-E053-6C04A8C09CA1}	1	2020-01-10	UB10 0JQ	F	N	L
5	{21E5FEB6-92B1-2439-E050-A8C06205342E}	100	2014-04-08	SW8 4DS	O	N	L
6	{21E5FEB6-F779-2439-E050-A8C06205342E}	100	2014-09-11	KT1 2DA	O	N	L
7	{21E5FEB6-E2DC-2439-E050-A8C06205342E}	100	2014-11-05	W1G 9XF	F	N	L
8	{42A5A70A-0A89-56E8-E050-A8C06205331E}	100	2016-06-10	N7 6JT	O	N	L
9	{4C4EE000-503D-1854-E050-A8C063054F34}	100	2016-10-04	SW1W 8TR	O	N	L
10	{3E0330F0-B134-8D89-E050-A8C062052140}	100	2016-08-24	IG11 8ED	O	N	L
11	{50F18103-CC4E-9FD5-E050-A8C063054923}	100	2016-06-23	E14 3AJ	O	N	F
12	{47844C80-C75E-8986-E050-A8C063056488}	100	2016-12-08	RM6 4BE	O	N	L
13	{75050A86-1E44-9A88-E053-6B04A8C02390}	100	2016-05-16	KT3 4NW	O	N	L
14	{49B7852A-E470-7921-E050-A8C063056E8D}	100	2016-10-19	SE1 6BN	O	N	L
15	{64342BFF-6936-422C-E053-6C04A8C0FB8A}	100	2017-08-08	SE9 2BG	O	N	L
16	{4C4EE000-326C-1854-E050-A8C063054F34}	100	2017-02-27	N17 9TA	O	N	L
17	{4C4EE000-3EB8-1854-E050-A8C063054F34}	100	2017-03-13	E17 3JX	O	N	L
18	{49B7852A-E00A-7921-E050-A8C063056E8D}	100	2017-01-27	BR1 2RG	O	N	L
19	{50F18103-CBE0-9FD5-E050-A8C063054923}	100	2017-05-05	HA8 9BX	O	N	L
20	{BC8936BC-5693-0E2C-E053-6C04A8C0DBF4}	100	2017-04-28	E6 1HW	O	N	L
21	{5376B386-3C0D-34C1-E053-6B04A8C09FF6}	100	2017-05-31	N17 0UR	O	N	L
22	{55BDCAE6-ECDD-521D-E053-6B04A8C0DD7A}	100	2017-06-02	SE9 6UD	O	N	L
23	{55BDCAE6-E8C3-521D-E053-6B04A8C0DD7A}	100	2017-04-11	CR0 2ER	O	N	L
24	{68FEB20C-401F-38DA-E053-6C04A8C051AE}	100	2017-12-12	N9 9JJ	O	N	L
25	{68FEB20C-5D98-38DA-E053-6C04A8C051AE}	100	2017-11-16	W1K 2HZ	O	N	F

Plotting the Points

Now go to Sheet 1. You will use several tools to explore the data. Start by dragging the [Price] measure to the Rows shelf.
In the top menu, click on /Analysis/Aggregate Measures to turn it off. Now each data point is represented by a circle. Please note that as this is a rather big dataset, going from aggregate measures (sum, mean, median, etc) to individual data points is likely to make your computer take a LONG time.

There are many overlapping points. Use transparency and reduce the size of the circles in the “Marks” cards. To do this Click on “Size” in Marks card and reduce the size to minimize the overlap. You can also use “Color” to reduce opacity.

Visualizing the data this way provides an easy way to check for outliers. You can now clearly see the property prices in extreme ends. Tableau also allows you to take a deeper look into the underlying data. Right click on one of the extreme data points then click

View Data and then Full Data. What do you observe? (See below if you don't see any data) Does this look like a standard property transaction? How can we use this information going forward?

5.1 Boxplots (a.k.a. Box-and-whisker plots)

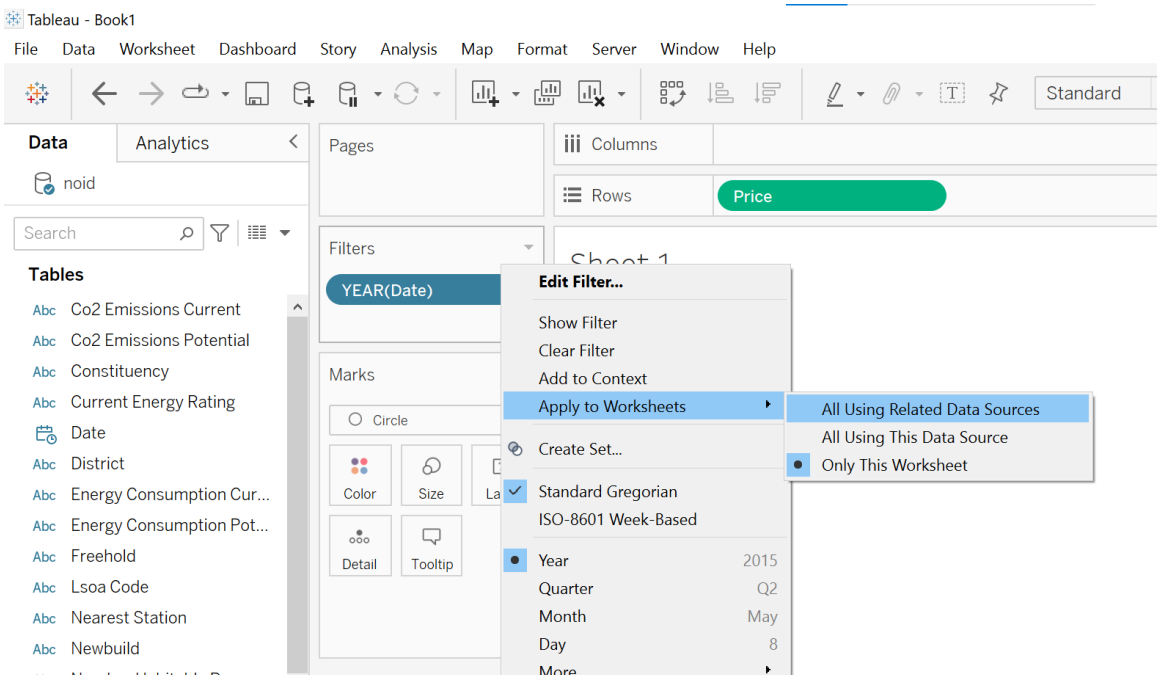
Next find the mean and median house prices in London and investigate if the prices are skewed.

The analysis in the previous section allowed us to detect outliers and extreme values. However, it does not provide much information about the ***distribution*** of prices.

Add a new sheet by click on "New Worksheet" that appears next to Sheet 1 and go to Sheet 2. Then drop Price to Rows Shelf. Again, make sure you go to **Analysis/Aggregate Measures** to turn off the automatic summation of all prices.

Add a box and whisker plot by choosing it from {Show Me} tab.

To simplify the analysis only consider sales in 2017 to 2020. To do this drop Date to "Filters", choose Years then 2017, 2018, 2019 and 2020. To make sure the same filters apply in all the worksheets we develop below, you can right click on the pill and choose "Apply to Worksheets" and "All using related Data Sources".



Because data is skewed it is not easy to visualize. Change the scale of the price axis by right clicking on this axis and then choosing “Edit axis” and then scale “logarithmic”. What do you observe?

Now drop Type 1 variable into column shelf. How does the distribution of the prices look different for two different types?

The Boxplot, invented by John Tukey (Princeton), is often used to describe the distribution of data. The key elements are:

- A box from the first quartile to the third quartile, with a division at the median is used to represent the center of the data. The length of the box is called the Inter Quartile Range (IQR).
- A line extends from the box to an upper and a lower whisker.
- The lower whisker is located at the minimum of the data, but no farther than 1.5 IQR below the first quartile.
- The upper whisker is located at the maximum of the data, but no farther than 1.5 IQR above the third quartile.
- Data outside the whiskers are considered outliers.

There are many outliers as you observed earlier. If you take a closer look, we see that the “property type” of these properties are listed as “B”. This is where you need to be careful because we are mainly interested in residential properties and these extreme data points are not for residential properties and they may have a huge impact on our analysis. For example, sale of LBS or of West Ham United Football Club’s stadium, Boleyn Ground, are also included in this list, which obviously are not residential properties.

What are the different *Type1* values?
A = Standard Price Paid entry, includes single residential property sold for value.
B = Additional Price Paid entry including transfers under a power of sale/repossessions, buy-to-lets (where they can be identified by a Mortgage), transfers to non-private individuals and sales where the property type is classed as ‘Other’.

On the lower end of the price spectrum, you will find data on sales on garages and storage space. To avoid the potential impact of these data on our analysis we can filter them out. Drop “Type 1” in Filters shelf and choose “A”. To have quick access to a filter you can show them on the right-hand side of the plot area by clicking on the pill dropped to filter and choosing Show Filter. Also do not forget to apply the filters to all sheets by right clicking on “Type 1” pill on the filter pane and choose “Apply to Worksheets” and “All using related data source” as described above.

What do you observe when you apply filters? How can the outliers affect our analysis if we do not filter out these transactions? We will take a deeper look into this issue in the next steps.

Turn the /Analysis/Aggregate Measures back on before you move on.

5.1 Histograms

Investigate the distribution of the house prices. Does the price distribution follow a normal distribution?

Although box-and-whisker plots are useful in detecting outliers and identify the mean and quantiles of the price data, it is difficult to read the actual distribution of the prices. Next you will use histograms.

Add a new worksheet and go to the newly created worksheet. Then drop “Price” to Rows and click on “Histogram” in the {Show Me} palette.

It will create a new dimension Price (bin) put in the {Column} shelf. That combined with the CNT(Price) in the {Rows} shelf generates the histogram.

A majority of homes are sold for less than 10 million. You can have a better visualization of the distribution of the prices by dropping “Price” to filter and filtering out those that were sold for more than 10 million.

*Default bin size is too large due to outliers in the data. You can change this by right clicking on **Price (bin)** in the “Tables” pane to the left (not the one in {Rows}) and choose edit. This brings up a window to select the parameters. Set “Size of Bins” to 100,000*

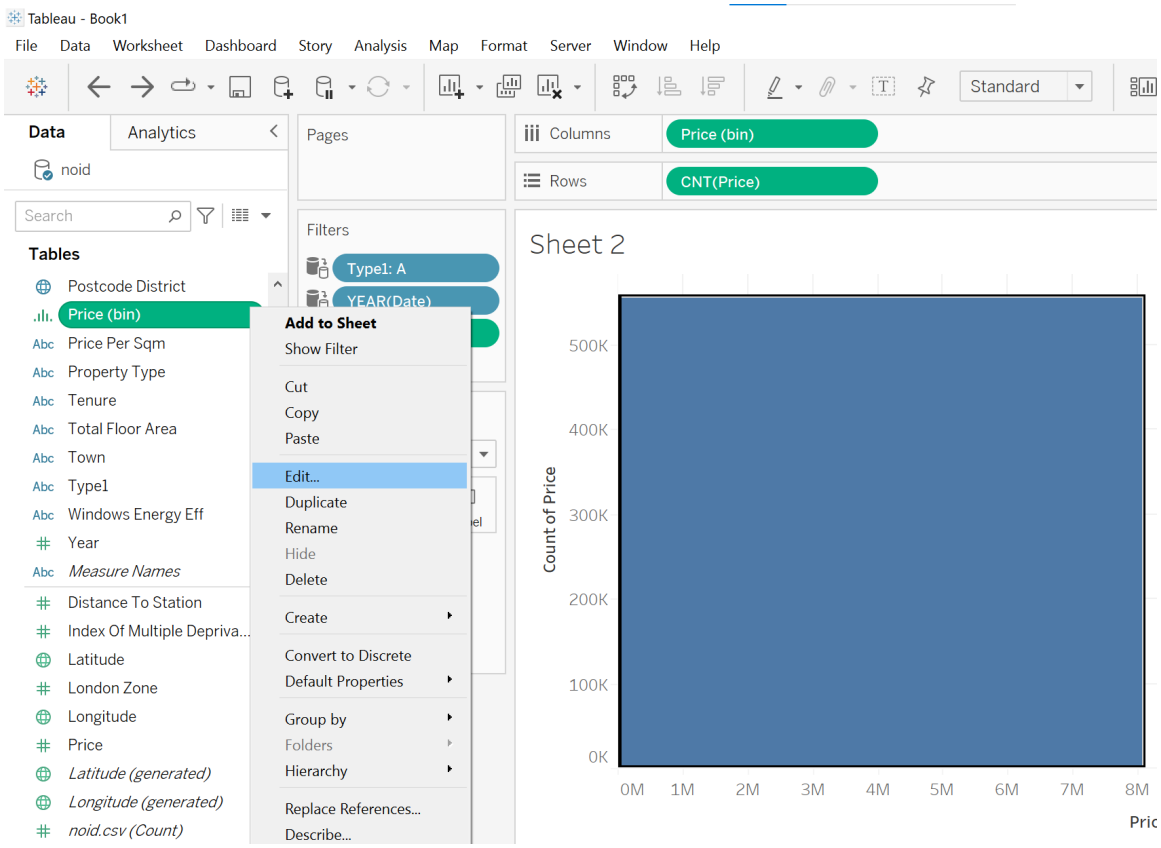
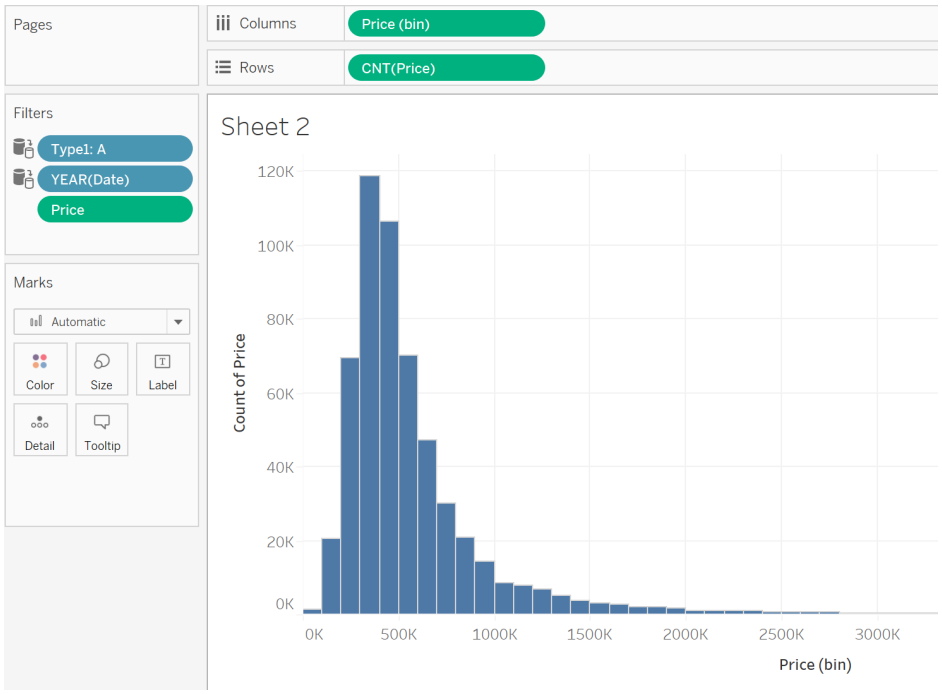


Tableau - Book1
File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Standard
Data
Analytics
noid
Search
Tables
Postcode District
Price (bin)
Price Per Sqm
Property Type
Tenure
Total Floor Area
Town
Type1
Windows Energy Eff
Year
Measure Names
Distance To Station
Index Of Multiple Depriva...
Latitude
London Zone
Longitude
Price
Latitude (generated)
Longitude (generated)
noid.csv (Count)
Add to Sheet
Show Filter
Cut
Copy
Paste
Edit...
Duplicate
Rename
Hide
Delete
Create
Convert to Discrete
Default Properties
Group by
Folders
Hierarchy
Replace References...
Describe...
Pages
Columns
Price (bin)
Rows
CNT(Price)
Filters
Type1: A
YEAR(Date)
Sheet 2
Count of Price
500K
400K
300K
200K
100K
0K
0M 1M 2M 3M 4M 5M 6M 7M 8M
Pric


Pages
Columns
Price (bin)
Rows
CNT(Price)
Filters
Type1: A
YEAR(Date)
Price
Marks
Automatic
Color
Size
Label
Detail
Tooltip
Sheet 2
Count of Price
120K
100K
80K
60K
40K
20K
0K
0K 500K 1000K 1500K 2000K 2500K 3000K
Price (bin)

What can you deduce from the histogram? Does it look like a bell-shaped curve? Is it right or left skewed? What's the implication of skewness?

Box-and-whisker plots are typically used in the initial steps of data exploration. Also, they are useful for comparisons, e.g., house prices in different boroughs (more on this below). Histograms provide more information about the distribution of the data, but they are not very effective in making comparisons, i.e., comparing the price distribution for different types of properties.

6. House Prices over the years

Investigate if London house prices are falling after Brexit and if Brexit had a different impact in different boroughs of London.

6.1. One Measure and One Nominal Dimension

You can do this in several ways in Tableau. Let's start with the simplest. First in your "Years" filter add years from 2007 to 2020. Then create a new worksheet.

Drag [Date] to the {Columns} shelf and [Price] to the {Rows} shelf to get the figure below and then choose "Horizontal Bars" from "Show Me" tab. If "Price" pill appears on Columns instead of Rows, you can switch the layout by clicking  on the menu bar. Also, you can change the summary statistic from sum to others, eg. mean, median, count, and etc, by clicking on the Sum(Price) pill and moving the pointer to Measure(Sum).

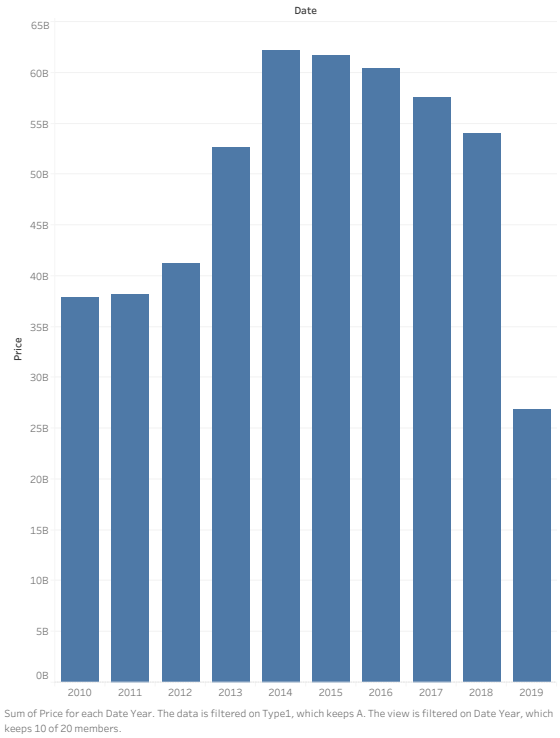


Tableau allows "Date" type variables to be treated as nominal or ordinal variables and which one you should use depends on your visualization goal. For example, if you are interested in visualizing the seasonality in home prices, you might want to treat date as nominal and plot prices vs moths. This way Tableau will find the average price in each month for the time period in the data. You will explore more differences in the next steps and one of the strengths of Tableau is that it allows you to switch seamlessly between these two.

Investigate how prices changed in quarters following Brexit.

You can do this by clicking on the plus sign next to the Year(Date) pill.
Naturally there is seasonality in the data; people are less likely to move in winter. To compare quarters in different years we can treat data as a categorical variable.

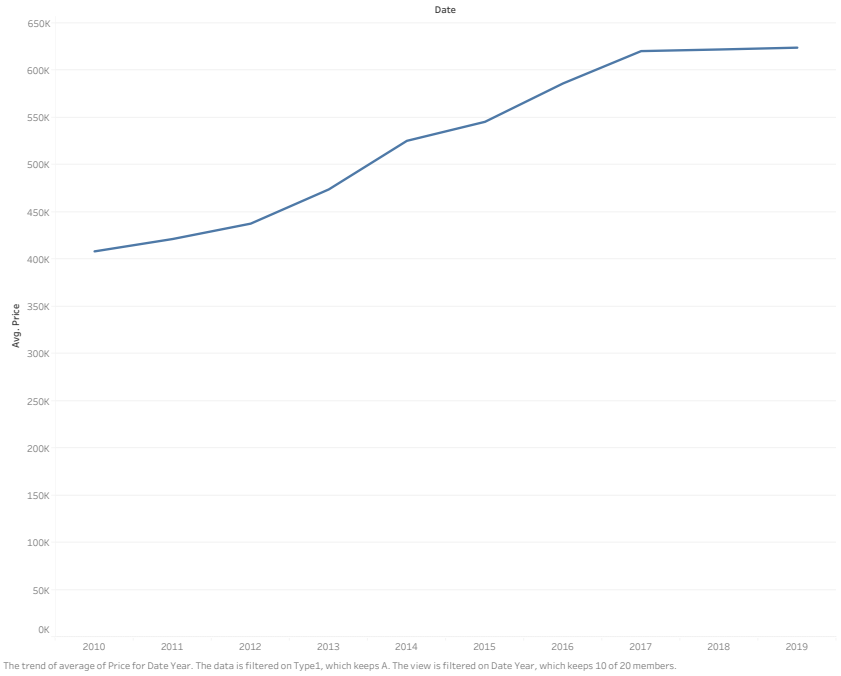
To do this make sure you have Year(Date) and Quarter(Date) in your columns shelve. Now change the order of Quarter and Year by dragging one of them. Compare the number of sales across the years in different quarters.

6.2 One Measure and One Ordinal Dimension

In the previous section you treated date as a categorical variable. It is actually easier to visualize house prices over the years by treating it as an ordinal variable. This will allow you to use line plots.

Let's now treat Date as an ordinal dimension. Instead of bar charts we will use line charts. Add a new worksheet. (Don't forget to save your work.)

Drag [Date] to the {Columns} shelf and [Price] to the {Rows} shelf. Click on the Date in the Column shelf and the second "Year" in the drop-down menu. This lays out the data linearly as opposed to aggregating over the months. In other words, it treats data in a linear manner instead of treating it as a categorical variable as we did in the previous part. Choose "Lines" from "Show Me" Card if it is not automatically selected. Display average prices over time by clicking on the Sum(Price) and choosing average, as you did before. It will produce a graph that is similar (perhaps not exactly the same) to the one shown below.



You can now see the value in filtering out unrepresentative data. Now add data for Type 1 "B" transactions by clicking on "Type 1" pill under Filters and choosing "Show Filter". Do you see a change in the trend? What does this imply?

Whether it's treated as a nominal or an ordinal variable, you can use similar tools to add more categorical variables to your plots. For example, you can easily switch from year to quarter as the unit of time by clicking on the + sign next to the date pill. Next drop a Price pill on "Detail" and choose Count(Price). Place the pointer on the plot to see how many houses are sold.

How many houses are sold in 2018? What does this imply for comparing average prices in 2017 and 2018?

To illustrate the use of "color", drop "Freehold" to Color in Marks Tab. Now try to drop another Price pill to Rows and view number of sales over time. You can display them in the same figure by choosing "Dual lines" from Show Me card.

6.3 Trend lines

Let's add some analytical tools to our visualization to explore the trends. Add a new worksheet to your workbook and plot average price vs year as in the last step, then

- Select Analysis/Trend Lines/Show Trend Lines from top menu.
- Right click on the trend line to bring up the options. Click on "Edit Trend Line" and then check the box "Show Confidence Bands" .
- You can also click on Describe Trend line to get a report on the regression results.

Investigate by how much the average house prices increased every year since 2005 and 2010. How about median prices? Are there any differences between how average and median prices changed?

Now investigate how the number of sales changed over the years on a new sheet, without including the data from 2021 (because the data for 2021 is not yet complete). What is the trend in number of sales?

Also visualize sales per quarter and months. You should observe an extreme outlier in the first quarter of 2016. What might have caused this outlier? What is the implication of this outlier on the number of sales in subsequent periods?

7 Advanced tools

Tableau provides a variety of advanced tools that can help visualize data better in different situations. You will explore a few of these tools in the following steps.

7. Additional tableau tools

Tableau provides a variety of advanced tools that can help visualize data better in different situations. I demonstrate a few in the following steps.

7.1. Percentage change

What is the annual rate of change in prices and sales volume in London house market?

This information is not included in the data provided but you can easily calculate it using a standard operation in Tableau. Although Tableau is not designed for data manipulation, it allows for basic calculations. Let's look at one of the most useful ones. See screencast 1

Plot average price vs year on a new worksheet. Right click on "Avg(Price)" pill and choose

"Quick Table Calculation" and then "Percent Difference". You can change the reference year by right clicking on the pill again and choosing "Relative to".

Is the average house price in London "slumping"? Is the sales volume decreasing? Use trend line and confidence bands to improve your visualization.

A more accurate assessment of price changes over the years would be plotting the median price per square meter (why?). We have data for the total floor area in square meters for about 70% of the data.

7.2. Map visualization

One of the data types in Tableau is Geographic, which Tableau uses for visualizing data on maps. We will use it to see average house prices in different locations on a London map. See screencast 2 for more details

Use a box plot to plot average price vs. postcode district on a new worksheet. Then choose one of the map options from *Show Me* card. To make sure that Tableau recognizes that the data is from the UK, choose Map (from top menu), Edit Locations, and choose United Kingdom for country. Do the results on the map make sense? You can improve the map visualization result by clicking on Color in the Marks card and then Edit Colors->Advanced.

To improve map visualization we can tell Tableau that London Zone is a geographical location variable. To do this click on the arrow next to *London Zone*, and choose Geographical Role->Create From->Postcode District as shown below.

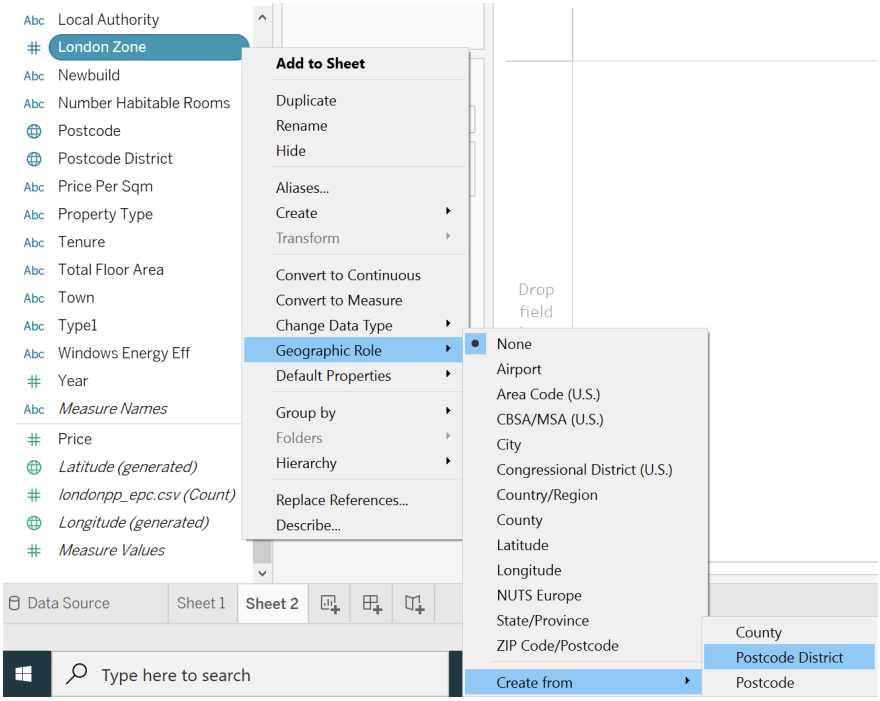


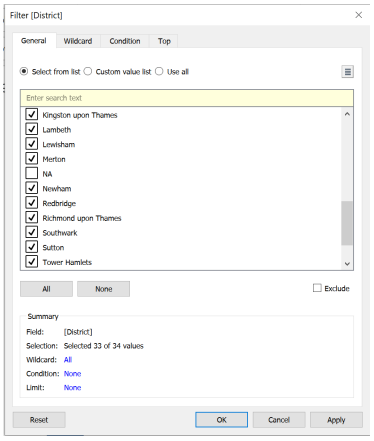
Figure 2 Assigning a geographical role to a variable

Use a box plot to plot average price by district and London zone on a new worksheet. Then choose one of the map options from Show Me. To make sure that Tableau recognizes that the data is from the UK, choose Map (from top menu), Edit Locations, and choose United Kingdom for country. Do the results on the map make sense? (Eg. check if central locations are more expensive.)

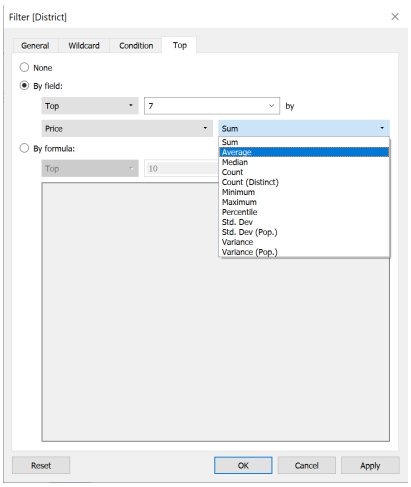
You can manipulate your geospatial visualisation to show % changes in each district, either as a year-on-year change, or relative to the first map you have—see screencast 3 for more details.

7.3. Filtering – basic interactivity

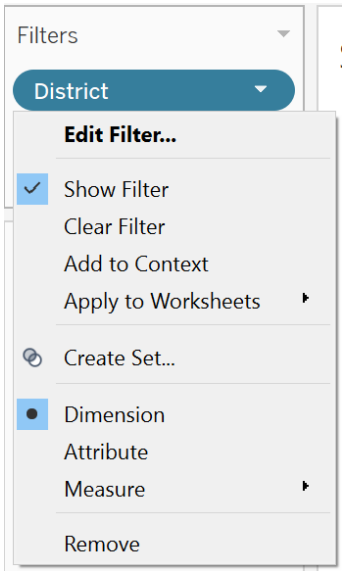
You can filter your data by dropping the variable you want to filter for in the Filter card. This is a quick way to filter out any NAs you may have in your dataset.



Besides selecting values of interest, you can easily and quickly choose the top ***k*** districts in terms of, say, average price.



Also, if you want to add basic interactivity and allow the user to choose, e.g., once you have added the variable you want to filter for in the Filter box, you select it and click on *Show Filter* – see screencast 4 for details



8. Story telling with Visualizations

Tableau has additional tools to summarize your findings to different audiences. In this part we will explore "stories". The other useful tool is dashboards, which provides at-a-glance views of key performance indicators.

- This screencast demonstrates how to make dashboards.

In this part you will first build a story and then post it on the internet on Tableau Public server. In order to build a data story, you need a story first and the visualizations to make the case for your story. Assuming you already built a series of visualizations, this screencast demonstrates how to make data stories using Tableau.

This document explains when and how to use dashboards and stories (read it after the workshop, it will be helpful for the assignment.)

9. Your task (i.e. assignment)

Can you identify a story using this rich data set? As discussed in class you should try to answer interesting business questions via visualizations. You need to submit your story as a URL link from *Tableau Public.* You can sign up in Tableau public here

https://public.tableau.com/desktop/signup_unification.html and then upload your work on Tableau Public, as detailed here https://www.thetableaustudentguide.com/tableau-public/publishing-to-tableau-public .

You can tell a story in 4-5 slides, there's no need to produce many visualizations. (See grading rubric under the assignment for more details.)

To get you started here are some questions that you might find interesting. Please do not feel obliged to use one of them. This is such a rich data set; you can come up with many different questions.

- How did Brexit (approved on a June 2016 referendum) affect London house prices? Note that even though the referendum took place in June 2016, the UK didn't formally leave the EU until 31 Dec 2020.
- How did Covid Pandemic affect London house prices?
- How did financial crisis of 2008 affect London house prices?
- Where is London growing towards?
- Which neighborhoods are financially sound for buying a property?