

TURF



Table of Contents

Overview and Background.....	4
Data requirements and preparation.....	5
Data used in TURF studies.....	6
Missing data.....	7
Sample size	7
TURF analysis.....	8
Bubble gum flavor preferences case study	9
TURF	9
The F in TURF	10
Constraining which alternatives appear	11
Understanding substitution using a duplication matrix.....	12
Using multiple correspondence analysis to understand substitution	12
Venn diagrams.....	13
Interactive Venn Diagrams	14
Upset plots.....	14
Using principal component analysis to understand substitution	16
Portfolio size (how many products?).....	17
TURF with many alternatives.....	18
TURF in Display.....	20
Importing and setting up the data	21
Viewing the reach of the alternatives	21

Running the TURF	22
Comparing TURFs with different portfolio sizes	23
Duplication matrix.....	23
Correspondence analysis of a square table	24
Principal components analysis.....	24
Venn diagrams.....	24
Interactive Venn diagrams.....	25
Upset plot.....	25
TURF in Q.....	26
Importing and setting up the data	27
Viewing the reach of the alternatives	27
Running the TURF	28
Comparing TURFs with different portfolio sizes	29
Duplication matrix.....	29
Correspondence analysis of a square table	30
Principal component analysis	30
Venn diagrams.....	30
Upset plot.....	30

Overview and Background

TURF is an acronym for **T**otal **U**nduplicated **R**each and **F**requency. It is widely used to answer two questions:

1. If a company can make a lot of possible products, but can only support a small number of them, which should it make?
2. Which *media vehicles* should a company focus on when distributing content? For example, which advertising networks and media should it focus upon?

This book covers everything from data preparation through to data visualization and reporting.

Data requirements and preparation

A TURF study requires a data set that consists of:

- Data from at least 100 people, and ideally many more
- For each person, measurements of the appeal of three or more different alternatives, which are typically products or content. These measurements can be from rating scales, historic purchase data, MaxDiff, or conjoint utilities, or preference predictions.

The data preparation consists of converting each measurement to a binary variable, where a 1 indicates the person likes whatever is being measured a 0 indicates otherwise.

Data used in TURF studies

The classic data used in a TURF study contains multiple responses, setup in the *binary format*, where:

- Each row represents an observation (e.g., a person, household, or company).
- Each column represents an *alternative* (e.g., flavors liked, or websites visited regularly).
- A value of 1 indicates that the alternative is applicable for that observation, and a 0 that it is not (e.g., a 1 may indicate products liked, or websites visited in the past month).

	Flavors: Classic Bubble Gum	Flavors: Super-Strong Bubble Gum	Flavors: Grape	Flavors: Sour	Flavors: Orange	Flavors: Strawberry	Flavors: Cola	Flavors: Apple	Flavors: Chocolate	Flavors: Peach	Flavors: Watermelon
1	1	0	1	1	1	0	1	0	0	0	0
2	1	1	0	1	1	0	0	1	0	0	0
3	1	1	1	1	0	0	1	0	0	0	0
4	0	0	0	1	1	0	0	1	1	0	0
5	1	1	0	0	1	1	0	0	1	1	1
6	0	0	0	1	1	1	1	1	0	0	0
7	0	0	0	1	1	0	0	1	1	0	0
8	1	1	1	0	0	0	0	1	0	0	0
9	1	1	0	1	1	0	1	0	0	0	0
10	1	1	0	0	0	0	0	0	0	0	0
11	1	1	1	0	0	1	0	0	1	0	0
12	1	1	1	1	0	0	1	0	0	0	0
13	1	1	1	0	0	0	0	0	0	0	0
14	1	1	1	1	0	1	0	0	0	0	0
15	0	1	1	0	0	0	0	0	0	0	0
16	0	0	0	0	1	1	1	1	0	0	0

Rating scales (e.g., purchase intent)

Data from purchase intention and other rating scales are also commonly used in TURF studies. In order to use this data, it is necessary to first convert the data into binary format. For example, assigning a 1 to everybody that has said “Definitely will buy” and “Probably will buy” and a 0 to everybody else.

MaxDiff data

MaxDiff data can be used in TURF as well, but it is necessary to first convert it into a binary format. There are a number of ways of doing this:

- Highest ranked alternatives. The highest ranked alternatives for each observation are assigned a value of 1 and the others a value of 0.

- Threshold-based selection. Some threshold is specified (e.g., utilities above the respondent's average), and alternatives with a utility above this value are assigned a value of 1 and others a value of 0.
- Probability-based rules, such as alternatives with a probability above 10% are given a value of 1. This last approach is inadvisable, as *scale factors* mean that the key determinant of probabilities is the amount of noise in their data, so if you use an approach like this you end up biasing the MaxDiff towards people that were indecisive and inconsistent in their MaxDiff.

Conjoint data

Conjoint data can also be used for MaxDiff, in pretty much the same way as with MaxDiff, using either levels of an attribute, or, simulated products.

Missing data

TURF does not work with missing data. And, imputation tends to be highly problematic. Non-variance-preserving imputation methods, such as replicating missing values with means or predictions from linear regression models, do not work as the resulting data is no longer binary. TURF can be used where the imputation is stochastic (i.e., imputes a 1 or a 0 probabilistically), but the resulting TURF is problematic, as the more noise in the imputation process the less reliable the TURF.

Sample size

TURF is an extremely greedy procedure in terms of sample size. As we will see in the next chapter, the differences between different portfolios tend to be very small and are typically not even close to being statistically significant. TURF studies with important strategic implications should have sample sizes in the thousands if possible. At the beginning of the chapter a minimum sample size of 100 was proposed. This is of course just an arbitrary line in the sand, but the important point to appreciate is that TURF is more sample size hungry than just about any other advanced technique.

TURF analysis

This chapter illustrates the mechanics of performing TURF using a case study of bubble gum flavors.

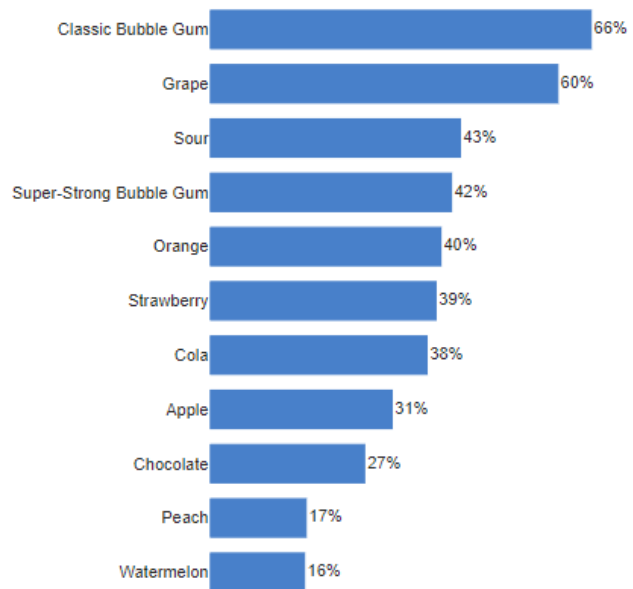
Bubble gum flavor preferences case study

A bubble gum manufacturer was selling two flavors of bubble gum – classic and Cola – and decided to extend the range to four flavors.

A survey of 712 parents was conducted, where the parents asked their children which of 11 flavors they liked. If you wish to reproduce the analyses in this book, you can download the data set from [here](#).

The percentages that like each of the flavors are shown to the right.

In the context of a TURF study, the “best four flavors” is defined as the four flavors that will maximize the proportion of people that like at least one of the flavors. This is known as *total unduplicated reach*, which is a term from media planning, where the focus was on understanding media vehicles (e.g., which combination of four magazines would reach the greatest number of people). It is usually just called “reach” for short.



TURF

The way that a basic TURF study works is that it examines the reach of all possible portfolios (combinations) of the four flavors (alternatives) and computes the reach of each of them. The results are shown to the right.

The winning combination is Classic Bubble Gum + Grape + Sour + Strawberry. This result is 95.1% - 94.4% = 0.7% above the number two portfolio, which has Cola appearing in-place of Grape. This 0.7% is not a big difference, representing only 5 of the 713 people. If you had to make a choice based on the data, you would prefer the first portfolio. But it would be a brave person to confidently assert that a 0.7% difference is meaningful.

Some people are disappointed when TURF fails to pick a clear winner. However, it is extremely good news, as it means that there is a lot of flexibility to take other factors into account (i.e., the research isn't closing off opportunities).

Total Unduplicated Reach and Frequency (TURF)

Top 10 portfolios with 4 alternatives

	Portfolio	Reach (%)	Frequency
1	Classic Bubble Gum, Grape, Sour, Strawberry	95.1	1,483
2	Classic Bubble Gum, Sour, Strawberry, Cola	94.4	1,322
3	Classic Bubble Gum, Grape, Strawberry, Cola	94.1	1,442
4	Classic Bubble Gum, Super-Strong Bubble Gum, Sour, Strawberry	94.0	1,352
5	Classic Bubble Gum, Grape, Orange, Strawberry	93.8	1,459
6	Classic Bubble Gum, Grape, Orange, Apple	93.7	1,405
7	Classic Bubble Gum, Grape, Orange, Chocolate	93.7	1,371
8	Classic Bubble Gum, Orange, Strawberry, Cola	93.7	1,298
9	Classic Bubble Gum, Grape, Orange, Cola	93.5	1,448
10	Classic Bubble Gum, Grape, Strawberry, Apple	93.5	1,399

n = 712 cases used in estimation; 330 portfolios examined out of a total of 330 possible portfolios (100%);

The F in TURF

The F in TURF refers to *frequencies*. The right-most column in the table above shows the frequencies. The value of 1,483 tells us that if we offer these four flavors, the total number of flavors that are liked across the total sample of 712 is 1,483 (i.e., on average, each of our 712 people liked fraction over two of the flavors in the portfolio).

Note that while the first two portfolios are very close in terms of reach, there is a big difference in terms of frequency, with the second alternative having a reach that is 11% lower. Looking at frequency makes the first, third, and fifth combinations seem the standouts.

Constraining which alternatives appear

The first portfolio is Classic + Grape + Sour + Strawberry. The third portfolio is Classic + Grape + Cola + Strawberry. So, the only real difference is Sour versus Cola. As the company currently makes cola, it is a “no brainer” to prefer portfolio three over either of the first two portfolios, as:

- The difference in results is marginal and could easily be a sampling fluke.
- It saves a lot of money to keep the existing flavors in the portfolio
- Cola is a known quantity which consumers are currently buying. The people that buy it may leave the market if it is gone, and there cannot be the same confidence that the small number of people that Sour would bring in will pan out as viable long-term consumers.

This suggests that we should only look at combinations that are constructed to always include the two current flavors of Cola and classic. The revised table is shown below. The two rows correspond to portfolios 2 and 3 from the earlier table.

Looking at the frequencies, which are again the bigger discriminator, we see strong evidence in favor of Classic + Grape + Strawberry + Cola, with Classic + Grape + Orange + Cola also doing reasonably well.

Total Unduplicated Reach and Frequency (TURF)

Top 10 portfolios with 4 alternatives

Portfolio	Reach (%)	Frequency
1 Classic Bubble Gum, Sour, Strawberry, Cola	94.4	1,322
2 Classic Bubble Gum, Grape, Strawberry, Cola	94.1	1,442
3 Classic Bubble Gum, Orange, Strawberry, Cola	93.7	1,298
4 Classic Bubble Gum, Grape, Orange, Cola	93.5	1,448
5 Classic Bubble Gum, Grape, Sour, Cola	92.3	1,472
6 Classic Bubble Gum, Strawberry, Cola, Apple	92.3	1,238
7 Classic Bubble Gum, Strawberry, Cola, Chocolate	92.1	1,204
8 Classic Bubble Gum, Orange, Cola, Apple	91.7	1,244
9 Classic Bubble Gum, Grape, Cola, Apple	91.3	1,388
10 Classic Bubble Gum, Orange, Cola, Chocolate	91.3	1,210

n = 712 cases used in estimation; 36 portfolios examined out of a total of 36 possible portfolios (100%) with the constraint(s) that all portfolios must include the alternative(s): Classic Bubble Gum, Cola; the number of possible unconstrained portfolios is 330;

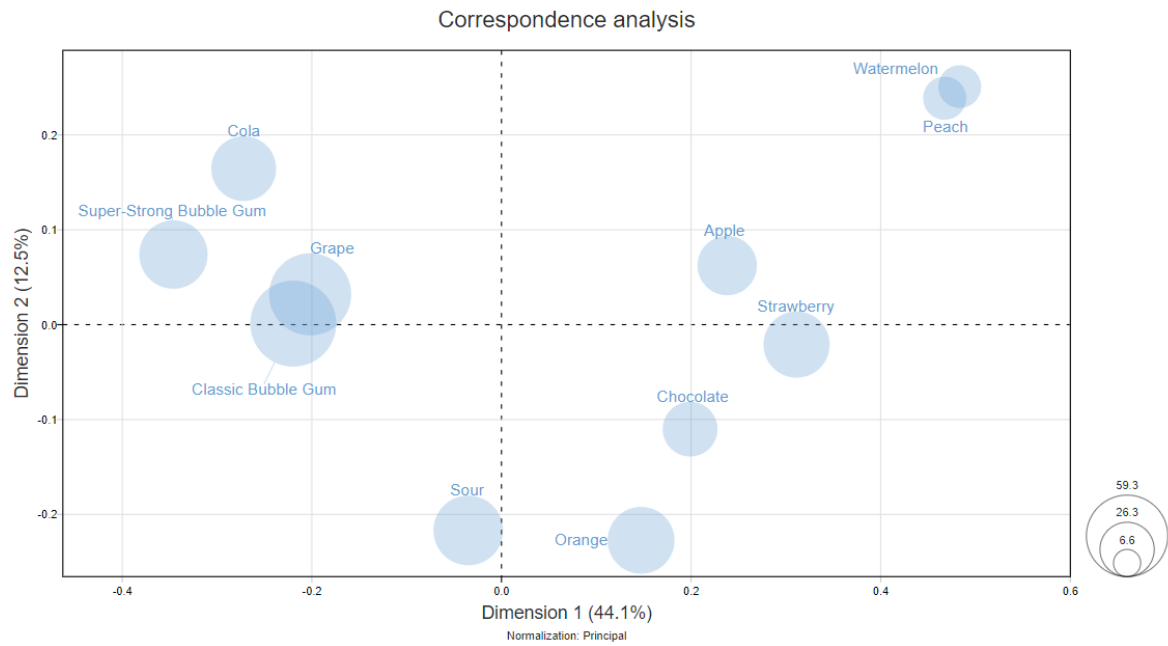
Understanding substitution using a duplication matrix

A *duplication matrix* is a crosstab showing the all the alternatives in both the rows and the columns. It is useful for understanding which alternatives are substitutes. The arrows below are showing where the substitution is statistically significant. For example, looking at the Classic Bubble Gum column, we can see that it is a strong substitute for Grape, Cola, and Super-strong, and very weak for the fruit flavors.

Total %	Classic Bubble Gum	Grape	Sour	Super-Strong Bubble Gum	Orange	Strawberry	Cola	Apple	Chocolate	Peach	Watermelon
Classic Bubble Gum	66% ↑	45% ↑	29%	35% ↑	25%	22% ↓	29% ↑	19% ↓	16%	8% ↓	9% ↓
Grape	45% ↑	60% ↑	26%	30% ↑	22% ↓	21% ↓	26% ↑	18%	15%	9%	9%
Sour	29%	26%	43% ↑	19%	25% ↑	16%	18%	16% ↑	13%	8%	8%
Super-Strong Bubble Gum	35% ↑	30% ↑	19%	42% ↑	15%	12% ↓	22% ↑	12%	9% ↓	6%	5% ↓
Orange	25%	22% ↓	25% ↑	15%	40% ↑	23% ↑	13%	16% ↑	12%	9% ↑	9% ↑
Strawberry	22% ↓	21% ↓	16%	12% ↓	23% ↑	39% ↑	12% ↓	16% ↑	14% ↑	13% ↑	11% ↑
Cola	29% ↑	26% ↑	18%	22% ↑	13%	12% ↓	38% ↑	12%	9%	7%	6%
Apple	19% ↓	18%	16% ↑	12%	16% ↑	16% ↑	12%	31% ↑	10%	10% ↑	10% ↑
Chocolate	16%	15%	13%	9% ↓	12%	14% ↑	9%	10%	27% ↑	7% ↑	6% ↑
Peach	8% ↓	9%	8%	6%	9% ↑	13% ↑	7%	10% ↑	7% ↑	17% ↑	7% ↑
Watermelon	9% ↓	9%	8%	5% ↓	9% ↑	11% ↑	6%	10% ↑	6% ↑	7% ↑	16% ↑

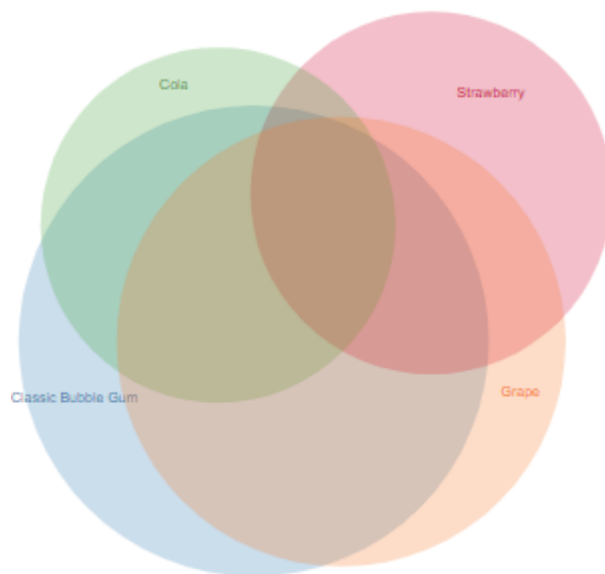
Using multiple correspondence analysis to understand substitution

A special variant of correspondence analysis, *known as correspondence analysis for square tables*, is particularly useful for visualizing the duplication table. The resulting *biplot* is shown below, with each bubble indicating the appeal of each alternative. The further apart any of the bubbles, the less they are substitutes. Ideally, we want to choose alternatives that are far apart. This raises problems with our current preferred portfolio of Classic + Grape + Strawberry + Cola, with Classic and Grape strongly overlapping, and Cola being in the same vicinity.



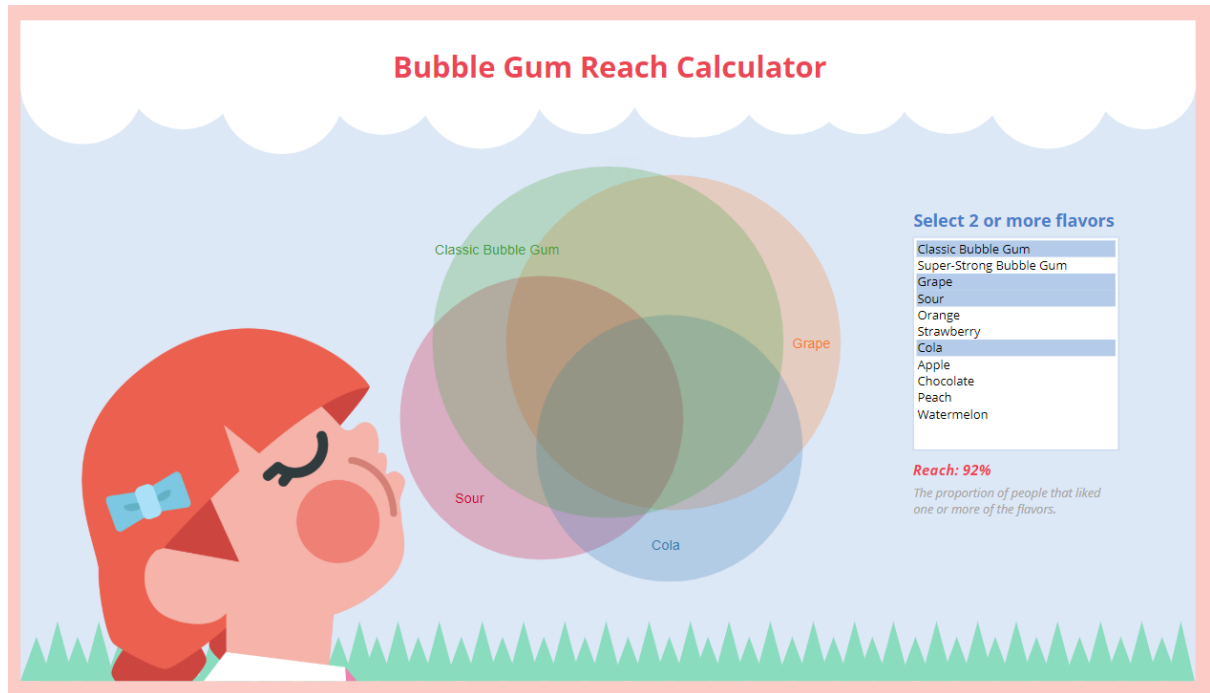
Venn diagrams

We can get a better understanding of the overlapping preference between the brands by using a *Venn diagram*. This diagram emphasizes that while Strawberry has the least overlap (as suggested by the correspondence analysis), but also that Grape and Classic have much higher overall appeal (i.e., bigger circles).



Interactive Venn Diagrams

Venn diagrams are often most useful when they are made interactive, so that you, or a stakeholder, can manually play around with different combinations. [Click here](#) to interact with this on a webpage.

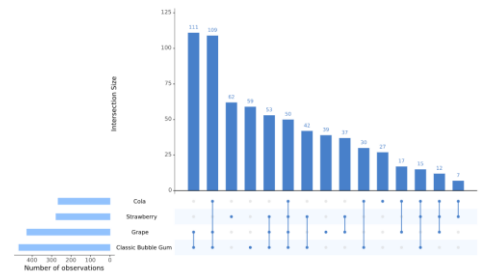


Upset plots

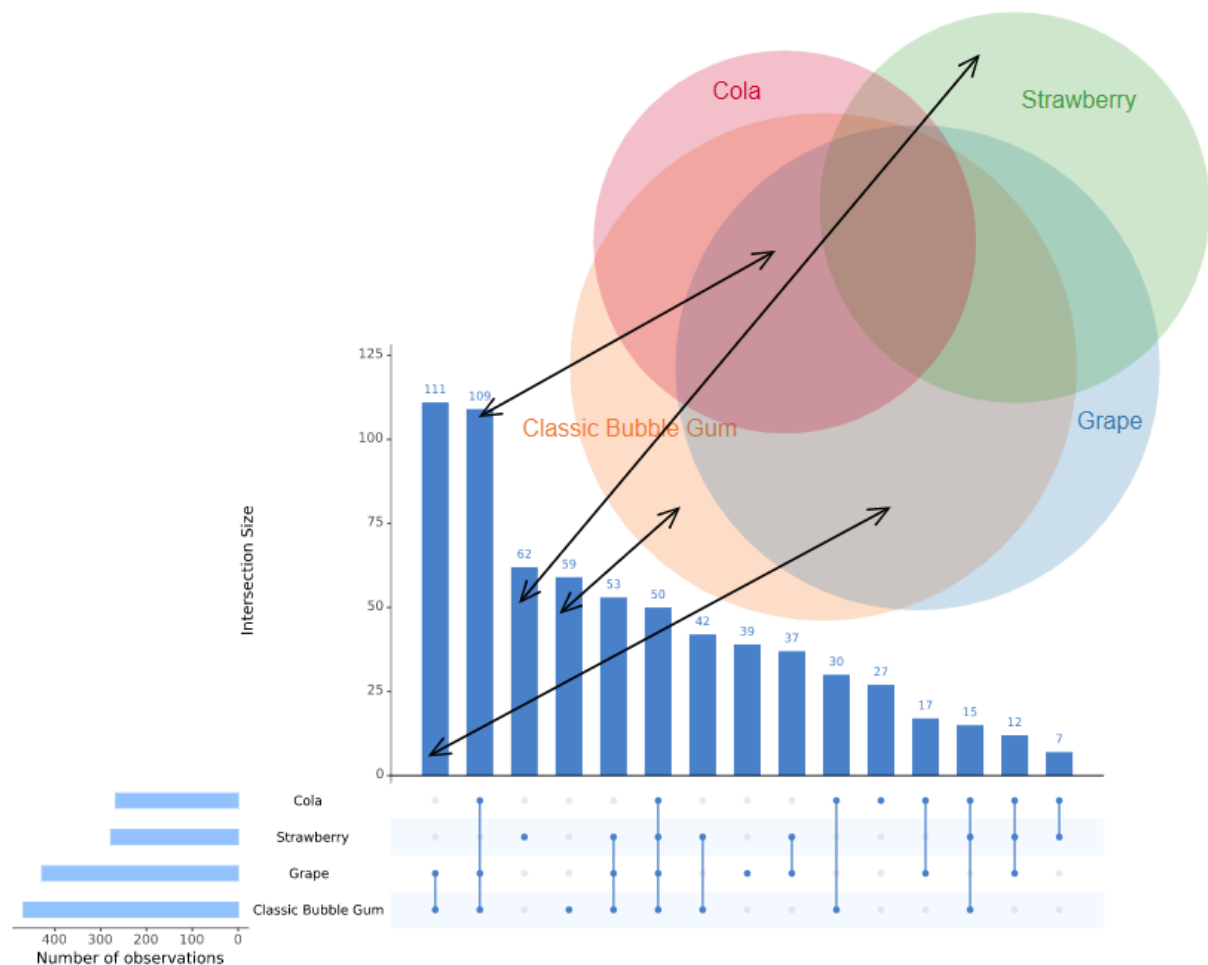
While Venn diagrams look cool, in practice they have a couple of limitations:

- They can misrepresent the overlap of the circles. Sometimes they show no overlap when there is some, and sometimes vice versa. This is because they are drawn in 2D, but often the overlap cannot be accurately shown in 2D.
- It's hard to label the high levels of overlap, which is why the diagram above shows no labels.

A solution to these problems is to use an *upset plot*. An example, too small to read, is shown to the right. Sadly, it is not nearly as pretty nor intuitive as a Venn diagram. It shows the same information as the Venn diagram attempts to represent but represents each the regions of the Venn diagram as a column and orders them according to size. The plot below shows that the most popular combination of the four flavors is Grape + Classic (the first column), with the next bar showing Grape + Cola + Classic.



The third column shows us Strawberry, which reinforces that it is doing the best job in terms of bringing extra 62 people into the market. By contrast, we can see that Grape and Cola bring in 39 and 27 respectively (these are each shown by columns with a single dark dot below the columns).



For all its ugliness, the upset plot is substantially more honest than the Venn diagram. Consider the intersection of Classic & Strawberry. The upset plot shows us that there is are 42 people that like only these two. But the Venn diagram show no intersection that only involves these two flavors! It's not because this Venn diagram is bad. Rather, it is because Venn diagrams can only show approximations for more than four or more flavors (Venn diagrams can be improved by changing the shapes from circles, but then they just become a different type of ugly to the upset plot).

If we focus on the intersections of grape, we can see that while Grape is a very popular flavor, its unique contribution is comparatively low, at 39 people. This is less than Strawberry, even though we know strawberry is less popular.

So, we have a choice: we can maximize frequency, which favors Grape, or go for variety, which says we should exclude Grape. The TURF table below shows the best 10 portfolios where we exclude Grape. The winning portfolio in this analysis is Classic + Sour + Strawberry + Cola.

Total Unduplicated Reach and Frequency (TURF)

Top 10 portfolios with 4 alternatives

Portfolio	Reach (%)	Frequency
1 Classic Bubble Gum, Sour, Strawberry, Cola	94.4	1,322
2 Classic Bubble Gum, Orange, Strawberry, Cola	93.7	1,298
3 Classic Bubble Gum, Strawberry, Cola, Apple	92.3	1,238
4 Classic Bubble Gum, Strawberry, Cola, Chocolate	92.1	1,204
5 Classic Bubble Gum, Orange, Cola, Apple	91.7	1,244
6 Classic Bubble Gum, Orange, Cola, Chocolate	91.3	1,210
7 Classic Bubble Gum, Super-Strong Bubble Gum, Strawberry, Cola	91.0	1,311
8 Classic Bubble Gum, Orange, Cola, Peach	90.7	1,138
9 Classic Bubble Gum, Sour, Cola, Apple	90.6	1,268
10 Classic Bubble Gum, Sour, Orange, Cola	90.4	1,328

n = 712 cases used in estimation; 28 portfolios examined out of a total of 28 possible portfolios (100%) with the constraint(s) that all portfolios must include the alternative(s): Classic Bubble Gum, Cola; the number of possible unconstrained portfolios is 210;

Using principal component analysis to understand substitution

Correspondence analysis provides a great summary of the data. But, it forces it into a two-dimensional plot, which means that it throws away some information. With larger numbers of alternatives, this becomes a serious problem.

We can get more granular understanding of substitution by using factor analysis or principal component analysis. The loadings for the data are shown below. It suggests that there are three basic groupings of flavors: Sour and Orange, the traditional flavors, and the fruits & chocolate flavors. This suggests that we should consider having one of Sour or Orange in our portfolio. As Sour is

further from Strawberry (see the correspondence analysis). This further supports the view that Classic + Sour + Strawberry + Cola is the best portfolio.

Principal Component Loadings: Flavors

3 components explaining 45.6% of the variance

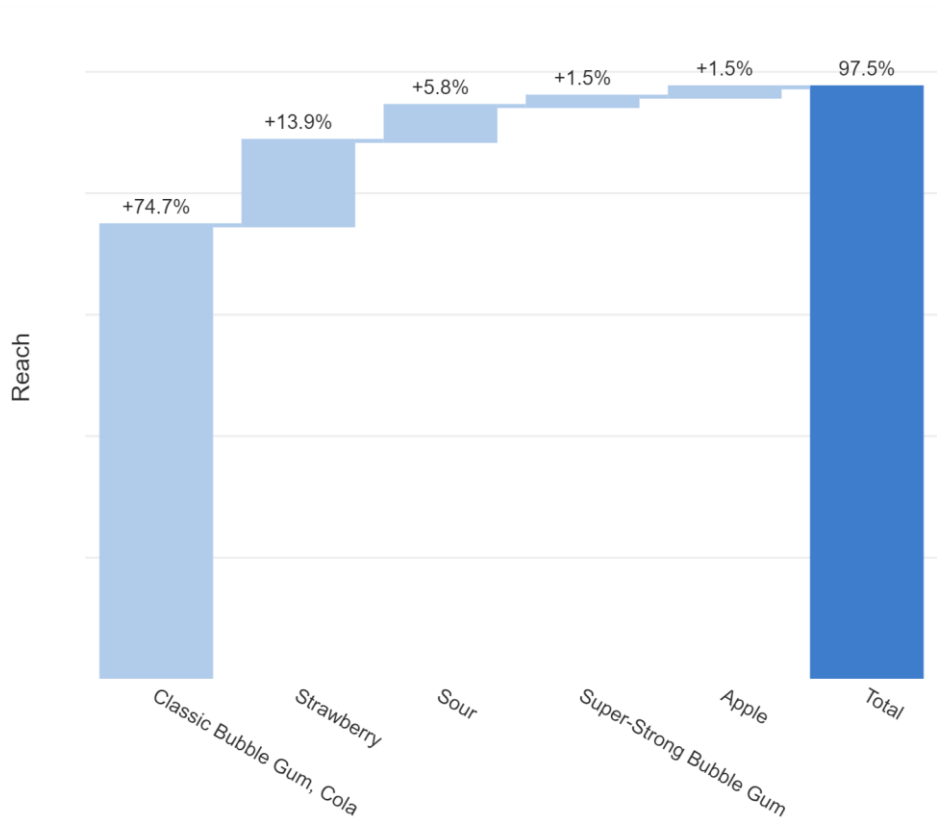
	Component 1 (18%) Eigenvalue: 1.97	Component 2 (15.6%) Eigenvalue: 1.72	Component 3 (12%) Eigenvalue: 1.32
Peach	0.753		
Watermelon	0.683		
Strawberry	0.641		
Apple	0.542		
Chocolate			
Super-Strong Bubble Gum		0.698	
Cola		0.644	
Classic Bubble Gum		0.635	
Grape		0.574	
Sour			0.831
Orange			0.746

Rotated loadings; Input: Correlation matrix; Missing data setting: Use partial data (pairwise correlations); Sample size: 712 to 712; Rotation: Varimax; Unrotated eigenvalues: (1) 2.29, (2) 1.53, (3) 1.18, (4) 0.97, (5) 0.92, (6) 0.84, (7) 0.77, (8) 0.72, (9) 0.67, (10) 0.61, (11) 0.50

Portfolio size (how many products?)

Should the bubble gum manufacturer offer 2, 3, 4, 5, or 6 different variants of gum? The way that this question is asked is to run TURF multiple times, and then create a *waterplot plot* like the one below which shows the incremental reach obtained through each additional alternative. As a rule, incremental gain in reach from each new product in the portfolio is much smaller than the one before.

Gut feeling would seem to suggest that three or four is likely the greatest number of products that make sense in this example, but ultimately this is a tradeoff that needs to consider the marginal costs of manufacturing, marketing, and distributing each extra product.



The incremental order will not always be consistent

As additional alternatives are added, it is often the case that all the previous alternatives remain. But this will not always happen. If you wish to have it so that it is always consistent, this can be done by constraining the TURF to always include the alternatives selected in the earlier TURF with fewer products.

TURF with many alternatives

This case study involves a comparatively small number of alternatives, with only 11 alternatives. It is not unknown for TURF studies to involve as many as 100 alternatives. However, this can present a problem, as the more alternatives, the longer it takes to evaluate them all.

There are only 330 possible combinations of four alternatives from 11 (if you remember your high school math, $330 = \binom{11}{4}$), so it was easy to evaluate them all in our case study. Evaluating all the combinations is sometimes known as the *exhaustive method*.

With 20 alternatives there are 4,845 combinations, with 50 there are 230,300, and with 100 there are 3.9 million, and you have to wait quite a while for a computer to evaluate all of them. For this reason, with larger number of alternatives, various shortcuts can be used that seek to find the best portfolios without evaluating all the possible combinations. These shortcuts are generally pretty good, but they aren't guaranteed to find all the good solutions.

The only other practical issue with TURF studies with many alternatives is that it gets a lot harder to get a good intuition about substitution. The trick to this tends to rely heavily on PCA.

TURF in Displayr

Importing and setting up the data

If you don't already have a data set loaded, start a new document in Displayr and press **Add a data set** (bottom-left corner). The data used in this book can be downloaded loaded from [here](#).

If you have a high-quality data file format, such as an SPSS file, the data to be used in the TURF will likely already be grouped together as a *variable set*. The data set used in this case study, however, is an Excel file, so we can save a lot of time in analysis by grouping together the binary variables that indicate flavor preferences. To tidy the data in the case study:

- Select from **Classic Bubble Gum** through to **Watermelon**
- Click **Combine** (from the top of the screen).
- In this example we can just press **Fix** and **OK** in the top right of the screen. In this step we are checking that we are counting the right values in the TURF. If our data is in some other format, such as numeric data, rating scales, or conjoint or MaxDiff data, instead select the variables, and then select **Properties > TRANSFORMATIONS > Binary Variables**, and finally click on one of the new variables and confirm that it is making the data binary in the appropriate way (see [Data requirements and preparation](#)).
- Give the variable set an appropriate label, in **Properties > GENERAL > Label** (top-right of the screen), such as `flavors`.

Viewing the reach of the alternatives

Drag the variable set onto a page. If everything has been setup correctly, you should have a table like the one to the right. Specific things to check for are:

- The percentages, which show the reach of each alternative, should be sensible.
- You can make your analysis a bit easier by sorting the categories (using drag and drop or by using the **Sort** menu).
- There should be no missing data. If there is any, it will be shown in the footer.

	%
Classic Bubble Gum	66% ↑
Super-Strong Bubble Gum	42% ▲
Grape	60% ↑
Sour	43% ▲
Orange	40%
Strawberry	39%
Cola	38%
Apple	31% ↓
Chocolate	27% ↓
Peach	17% ↓
Watermelon	16% ↓
NET	100% ↑

Flavors SUMMARY
sample size = 712; 95% confidence level

- If the NET is not 100% it means that there are some respondents who don't like any of the alternatives.

Running the TURF

To run the TURF:

1. Press **Anything > Advanced Analysis > TURF > TURF Analysis**
2. Drag the variables representing the alternatives into the **Alternatives** box (right-side of the screen)
3. Specify your desired **Portfolio size**.

Constraints

Rules regarding which alternatives should and should not appear are specified in the **CONSTRAINTS** options (right-side of the screen, in the middle).

Must include Alternatives selected in **Must include** are automatically included in all portfolios that are investigated.

Mutually exclusive alternative sets You can use the **Mutually exclusive set** options to specify sets of mutually alternatives. Only one of the alternatives selected in each set is permitted to be included the portfolios that are evaluated.

Minimum alternatives per case The minimum number of alternatives that a respondent (case) can like in order for the respondent to be included in the reach calculation. This is typically set to 1. Where the goal is to, for example, create an assortment (e.g., a bag of mixed candies), then higher values are appropriate.

Maximum alternatives per case The maximum number of alternatives that a respondent (case) can like in order for the respondent to be included in the reach calculation. This value needs to be less than the number of alternatives in order for it to have any effect (by default it is set to 1,000).

Setting this to low values, has two effects:

1. The overall reach will be decreased.
2. The portfolios that are constructed are made up of alternatives for which each, on their own, have comparatively small reach.

Advanced settings

Maximum portfolios to Investigate When the total number of possible combinations (ignoring constraints) exceeds this value, a Monte Carlo algorithm is run. You can work out which algorithm has been run by examining the footer of the outputs.

Number of portfolios to keep By default, the best 10 portfolios are shown in the outputs.

Comparing TURFs with different portfolio sizes

1. Create a TURF, as described above.
2. Copy it (using the **Duplicate** button) and change the **Portfolio size**.
3. Repeat these two steps until you have as many alternatives as you want.
4. **Anything > Advanced Analysis > TURF > Incrementality plot**
5. Select the **TURFs**, making sure you select them in order, starting with the smallest portfolio size.

Duplication matrix

To create a duplication matrix:

1. Drag the variable set onto the page so that it forms a summary table.
2. Drag it again, over the table that has been created, releasing it into the **Columns** box.
3. Change **Inputs > STATISTICS > Cells** from **Column %** to **Total %** (right-side of the page, when the table is selected).

Correspondence analysis of a square table

1. Click on a duplication matrix and press, on your keyboard, **Ctrl+C** for copy.
2. **Anything > Advanced Analysis > Dimension Reduction > Correspondence Analysis of a Square Table.**
3. Click into the **Input table** field and paste (**Ctrl-V**) the name of the duplication matrix into the field and then select it.
4. Change the **Output** to **Bubble Chart**.

Principal components analysis

1. **Anything > Advanced Analysis > Dimension Reduction > Principal Components Analysis.**
2. Drag across the variable set of alternatives and release it into **Variables**.

Venn diagrams

1. **Visualization > Venn Diagram**
2. Drag across the variable set from the **Data Sets** tree and release it on **Variables in 'Data'**. If you instead select each of the variables separately, hooking it up to be interactive, which is described below, won't work.
3. In **Inputs > COLUMN MANIPULATIONS > Columns to show** type the alternatives you want to show. If you type a lot, it will take a long time to compute (the math of Venn diagrams is surprisingly complicated with four or more alternatives). For example, type:
Classic Bubble Gum, Cola, Orange, Sour

Interactive Venn diagrams

Once you have performed the steps in [Venn diagrams](#):

1. **Anything > Page Design > Control > List Box**
2. Type the names of your alternatives into the **Item list**. Alternatively, leave the Item List empty, and instead use:
 - a. Click the **Calculation** button and drag a box on the page with your mouse
 - b. Enter the **R CODE**: `names(flavors)[-NCOL(flavors)]`, replacing `flavors` with the name of your variable set
 - c. Hide the R Output that contains the names of the alternatives using the **Hide** button
 - d. Click onto the list box and set **Control > Items from R Output** to the R Output that you just created
3. Set **Control > Selection Mode** to **Multiple selection** and choose, say, four of the alternatives.
4. Click on the Venn diagram and set **Inputs > COLUMN MANIPULATIONS > Select columns to show by** to **Choosing from Combo Box or List Box control**
5. Set **Columns to show** to the list control (probably called `List.box`).

Upset plot

1. **Anything > Advanced Analysis > TURF > Upset Plot**
2. Select the **Binary variables** to compare (i.e., the portfolio of interest).

TURF in Q

Importing and setting up the data

Open Q, start a new project, and click **File > Data Sets > Add to Project** to add your data set. The data used in this book can be downloaded loaded from [here](#).

If you have a high-quality data file format, such as an SPSS file, the data to be used in the TURF will likely already be grouped together as a *question*. The data set used in this case study, however, is an Excel file, so we can save a lot of time in analysis by grouping together the binary variables that indicate flavor preferences. To tidy the data in the case study:

- Go to the **Variables and Questions** tab.
- Select the variables representing from **Classic Bubble Gum** through to **Watermelon**
- Right-click on one of the row numbers and choose **Set Question**.
- Make sure the **Question Type** is set to **Pick Any**.
- Enter an appropriate Name, such as `flavors`.
- Click **OK**.

Viewing the reach of the alternatives

On the **Outputs** tab, in the blue dropdown menu, select the question (e.g., `flavors` if you are replicating the case study). Specific things to check for are:

- The percentages, which show the reach of each alternative, should be sensible.
- You can make your analysis a bit easier by sorting the categories (right-click on the top of the column of numbers and use **Sort by**).
- There should be no missing data. If there is any, it will be shown in the footer.
- If the NET is not 100% it means that there are some respondents who don't like any of the alternatives.

%	%
Classic Bubble Gum	66%↑
Super-Strong Bubble Gum	42%
Grape	60%↑
Sour	43%↑
Orange	40%
Strawberry	39%
Cola	38%
Apple	31%↓
Chocolate	27%↓
Peach	17%↓
Watermelon	16%↓
NET	100%↑

Running the TURF

To run the TURF:

1. Select **Automate > Browse Online Library > TURF > TURF Analysis**
2. Select the *question* representing the alternatives into the **Alternatives** box (right-side of the screen).
3. Specify your desired **Portfolio size**.

Constraints

Rules regarding which alternatives should and should not appear are specified in the **CONSTRAINTS** options (right-side of the screen, in the middle).

Must include Alternatives selected in **Must include** are automatically included in all portfolios that are investigated.

Mutually exclusive alternative sets You can use the **Mutually exclusive set** options to specify sets of mutually alternatives. Only one of the alternatives selected in each set is permitted to be included the portfolios that are evaluated.

Minimum alternatives per case The minimum number of alternatives that a respondent (case) can like in order for the respondent to be included in the reach calculation. This is typically set to 1. Where the goal is to, for example, create an assortment (e.g., a bag of mixed candies), then higher values are appropriate.

Maximum alternatives per case The maximum number of alternatives that a respondent (case) can like in order for the respondent to be included in the reach calculation. This value needs to be less than the number of alternatives in order for it to have any effect (by default it is set to 1,000).

Setting this to low values, has two effects:

1. The overall reach will be decreased.
2. The portfolios that are constructed are made up of alternatives for which each, on their own, has comparatively small reach.

Advanced settings

Maximum portfolios to Investigate When the total number of possible combinations (ignoring constraints) exceeds this value, a Monte Carlo algorithm is run. You can work out which algorithm has been run by examining the footer of the outputs.

Number of portfolios to keep By default, the best 10 portfolios are shown in the outputs.

Comparing TURFs with different portfolio sizes

1. Create a TURF as described above.
2. Copy and paste it (by right-clicking it in the **Report**) and change the **Portfolio size**.
3. Repeat these two steps until you have as many alternatives as you want.
4. **Automate > Browse Online Library > TURF > Incrementality plot.**
5. Select the TURFs, making sure you select them in order, starting with the smallest portfolio size.

Duplication matrix

To create a duplication matrix:

1. Create a table in the **Outputs** tab with your question (in this example, `flavors`) select in both the blue and brown drop-down menus
2. Right-click on the table, go into **Statistics – Cells** and add the **Total %**.
3. Again, go into **Statistics – Cells** and remove the **Column %**.

Correspondence analysis of a square table

1. Select **Create > Dimension Reduction > Correspondence Analysis of a Square Table**.
2. Click into the **Input table** menu on the right and select the table for your duplication matrix (created in the last step).
3. Change the **Output** to **Bubble Chart**.

Principal component analysis

1. Select **Create > Dimension Reduction > Principal Components Analysis**.
2. In **Variables** on the right, select the *question* that contains the options that you analyzed in your TURF (in this case, *flavors*).

Venn diagrams

1. **Create > Charts > Visualization > Venn Diagram**.
2. In the options on the right, click into the **Variables** box, and select the variables for the options that you want to include in the Venn Diagram.

Upset plot

1. **Automate > Browse Online Library > TURF > Upset Plot**
2. Select the **Binary variables** to compare (i.e., the portfolio of interest).

Want to cut your analysis and reporting time in half?

See Displayr in action →

Analysis and reporting software built to save you time

ML-powered text coding

Automatic, intuitive, stat testing

DIY dashboarding

Automatic data cleaning and tidying

Quickly create lots of crosstabs

Analyses that auto update with new data