Design of "Can the Success of Bachelor and Bachelorette Contestants Be Predicted?" Data Visualization

Introduction

This data visualization analyses the American TV shows *The Bachelor* and *The Bachelorette*. Over a 10-week period, contestants attend weekly dates and elimination ceremonies, culminating in the top four contestants achieving a hometown date from which an eventual winner is chosen. Two milestones over the course of the show are a first impression rose and a one-on-one date in the first four weeks. The supposed correlation of these milestones to an eventual win is a well-known assumption made by a large portion of the shows' audiences. This project can determine whether such a hypothesis is statistically justified. These graphs visualize how many contestants who received a one-on-one date in the first four weeks or a first impression rose on the first night, made it to the top four, how many of those top four went on to win and how many of the winners are still together with their partners. Both visualizations are overlapping grouped bar charts, comparing the two shows to each other.

Goals

- Simple layout that is thematically consistent

- Aesthetically related to the shows

- Clearly conveys the impact the two milestones have on succeeding on the show

- Smoothly transitions from one visualization to the next

- The visualizations work together

Data Analysis

The data used for these visualizations was extracted from a larger dataset containing information on 13 seasons of the Bachelorette and 21 seasons of the Bachelor. The dataset was compiled by FiveThirtyEight (https://github.com/fivethirtyeight/data/tree/master/bachelorette). The original dataset was cleaned and manipulated within Excel. For both the winners and one-on-dates, these figures spanned several columns. The numbers were calculated as shown in the screen shots below.

Screenshot 1

Cell N2: `=IFS(I:I="W","W",J:J="W","W",K:K="W","W",L:L="W","W",M:M="W","W",TRUE,"")`

C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	DATE
CONTESTA	ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	W	ST	DATES-1	DATES-2	DATES-3	DATE
ID	1	2	3	4	5	6	7	8	9	10			1	2	3	
13_BRYAN	R1		R	R			R			W	W	T			D6	D13
13_PETER_K	R					R	R			E				D1	D6	D13
13_ERIC_B		R				R	R		E					D10	D8	D13
13_DEAN_U	R		R			R	R	E						D8	D8	D1
13_ADAM_G							ED							D10	D8	D13
13_MATT_M							ED							D10		D13
13_ALEX_B		R			E									D10	D6	D13
13_WILL_G				R	ED									D10	D6	D13
13_KENNY_L				R	ED									D8	D8	D13
13_ANTHONY_B		R		E											D1	D13
13_JOSIAH_G	R			E										D10		D13
13_LEE_G				ED										D10	D8	D13
13_IGGY_R			E											D8		D13
13_JONATHAN_T			E											D8	D6	D13
13_JACK_S			ED											D8	D8	D1
13_BRADY_E		E														D8
13_BRYCE_P		E														D8
13_DIGGY_M		E												D10		
13_FRED_J		ED												D8	D6	
13_BLAKE_E	E													D8		
13_JAMEY_K	E													D10		
13_LUCAS_Y	E													D8		

Sheet tabs: **bachelorette** | Sheet1 | Sheet2 | Sheet8 | Sheet3 | R1 | Winners | Sheet5 | D1 | …
Status bar: Ready — 100%

Screenshot 2

Cell T2: `=IFS(Q:Q="D1","D1",R:R="D1","D1",S:S="D1","D1",TRUE,"")`

J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	DA
ELIMINATI	ELIMINATI	ELIMINATI	ELIMINATI	W	ST	DATES-1	DATES-2	DATES-3	DATES-4	D1	DATES-5	DATES-6	DATES-7	DATES-8	DATES-9	DA
7	8	9	10			1	2	3	4		5	6	7	8	9	
R			W	W	T			D6	D13		D1	D7	D1	D1	D1	D1
R			E				D1	D6	D13	D1	D9	D7	D1	D1	D1	D1
R		E					D10	D8	D13		D9	D1	D3	D1	D1	
R	E						D8	D8	D1	D1	D9	D7	D1	D1		
ED							D10	D8	D13		D9	D7	D3			
ED							D10		D13		D9	D7	D3			
							D10	D6	D13		D9	D7				
							D10	D6	D13		D9	D1				
							D8	D8	D13		D2	D7				
								D1	D13	D1	D9					
							D10		D13		D9					
							D10	D8	D13		D2					
							D8		D13							
							D8	D6	D13							
							D8	D8	D1	D1						
									D8							
									D8							
							D10									
							D8	D6								
							D8									
							D10									
							D8									

Sheet tabs: **bachelorette** | Sheet1 | Sheet2 | Sheet8 | Sheet3 | R1 | Winners | Sheet5 | D1 | …
Status bar: Select destination and press ENTER or choose Paste — 100%

These were then grouped; the data was filtered and the final computations are demonstrated in the screenshots below.
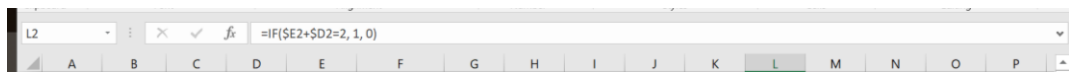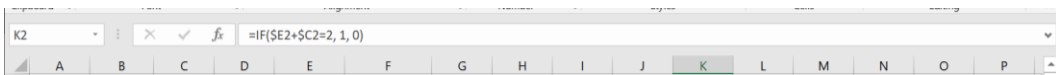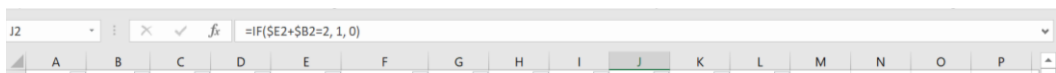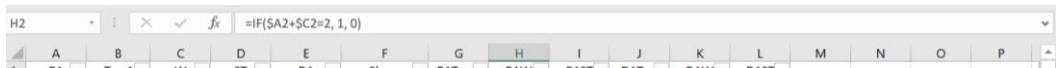
A2 | =IF(bachelorette!$D3="R1", 1, 0)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	R1	Top4	W	ST	D1	Show	R1Top	R1W	R1ST	D1Top	D1W	D1ST				
2	1	1	1	1		0 Bachelorette	1	1	1	0	0	0				
3	0	1	0	0		1 Bachelorette	0	0	0	1	0	0				
4	0	1	0	0		0 Bachelorette	0	0	0	0	0	0				
5	0	1	0	0		1 Bachelorette	0	0	0	1	0	0				
6	0	1	0	0		0 Bachelorette	0	0	0	0	0	0				
7	0	1	0	0		0 Bachelorette	0	0	0	0	0	0				
8	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
9	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
10	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
11	0	0	0	0		1 Bachelorette	0	0	0	0	0	0				
12	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
13	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
14	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
15	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
16	0	0	0	0		1 Bachelorette	0	0	0	0	0	0				
17	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
18	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
19	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
20	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
21	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
22	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
23	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				
24	0	0	0	0		0 Bachelorette	0	0	0	0	0	0				

bachelorette | Sheet1 | Sheet2 | Sheet8 | Sheet3 | R1 | Winners | Sheet5 | D1

Ready

B2 | =IF(ISERROR(MATCH("E*", bachelorette!$D3:$I3, 0)), 1, 0)

C2 | =IF(ISERROR(MATCH("W", bachelorette!$D3:$M3, 0)), 0, 1)

D2 | =IF(bachelorette!$O3="T",1, 0)

E2 | =IF(ISERROR(MATCH("D1", bachelorette!$P3:$S3, 0)), 0, 1)

G2   fx   =IF($A2+$B2=2, 1, 0)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P

H2   fx   =IF($A2+$C2=2, 1, 0)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P

fx   =IF($A2+$D2=2, 1, 0)

C	D	E	F	G	H	I	J	K	L	M	N	O	P

J2   fx   =IF($E2+$B2=2, 1, 0)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P

K2   fx   =IF($E2+$C2=2, 1, 0)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P

L2   fx   =IF($E2+$D2=2, 1, 0)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P

The First Impression Roses and the One-on-One Dates tables were created and graphed within excel.
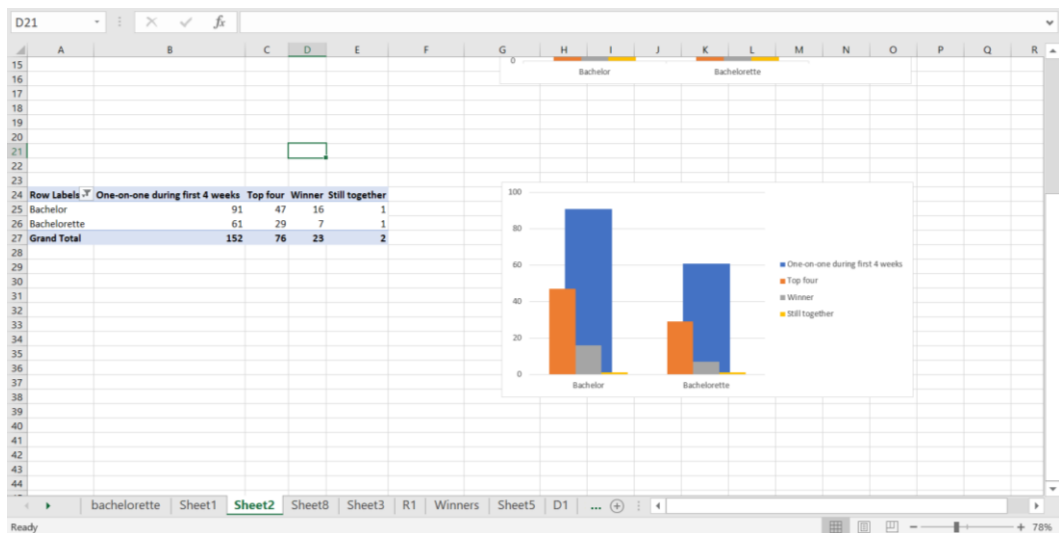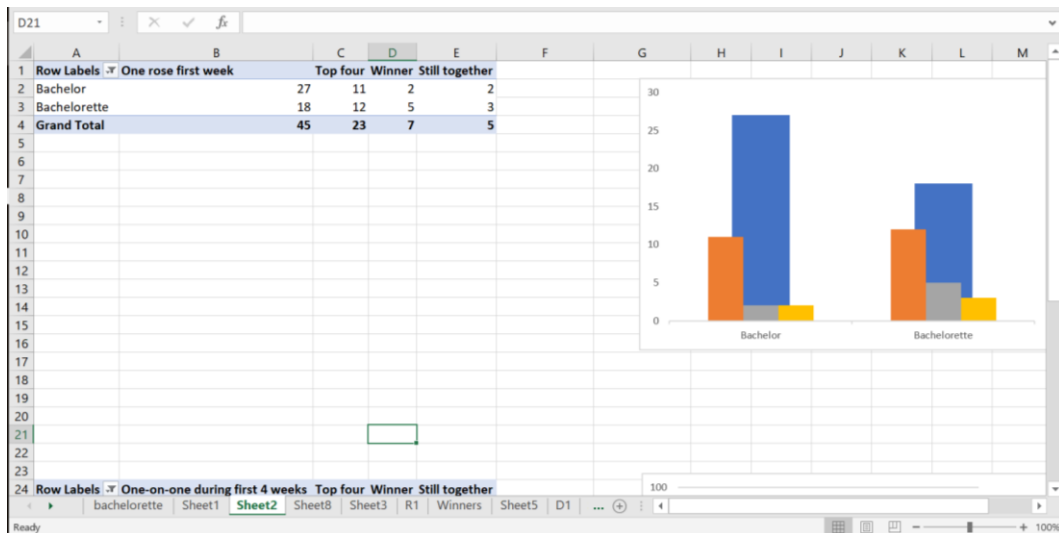
Then each table was saved as individual csv files and parsed into the JavaScript code.

Visualization Design

The design was created with the intention that the viewer would be able to visualise the significance of the first impression roses and one-on-one dates in the first four weeks. Despite utilising different data sets, the two visualizations were designed with a smooth transition so that they could work in sync as one visualization.

Development

The format of the visualization changed several times. Originally, it was an animated bar chart based on this (https://bl.ocks.org/1wheel/0a58f8bf5a14f6a534b9043a9c63dd1d). It then shifted towards a clustered-to-stacked bar chart. Eventually, after working with the data, the best option seemed to be an overlapping clustered bar chart because each variable (total, top four, winners, still together) are all interconnected. With a little added opacity to each bar, viewers are able to see how each variable relates to each other. However, rearranging the x-axis in D3 proved to be difficult, and the result was just a regular bar chart.

Final

The final design is simple and implements colours from the Bachelorette/Bachelor franchise. Colour scheme stays consistent across both visualizations in order to reinforce that they are connected. Data related to *The Bachelor* is shown in shades of red whilst *The Bachelorette* data is shown in shades of pink. The opening display consists of an empty axis, the visualization title and two buttons ('First Impression Rose' and 'One-on-One Date'). These buttons simplify the switch between the two visualizations. The uniform design throughout the visualization experience ensures that the data is the primary focus.

Evaluation and Conclusion

The visualization draws the conclusion that neither show milestone is particularly significant in making it to the top four, winning or staying together with the bachelor or bachelorette. It was designed for an audience that is familiar with the show structure as there is no additional explanation about the different elements or the timeline of the show.

There were a couple of difficult aspects writing the code for this visualization. Applying different colour schemes to clusters on the same x-axis was one of the more difficult elements to implement. In addition, it took some trial and error to load both csv files and ensuring that they worked cohesively and did not interfere with each other.

This visualization meets all the outlined goals. It is simple and effective in delivery of the data. The web app consists of one HTML page styled and positioned with CSS elements. As my D3 skills develop, future code could be streamlined and simpler.