

# DM & Visualization - Exam 2021-02-01 - Solution



Figure 1: Breakdown of wedding costs according to WeddingWire

## Analysis

Analyze the above graph reporting the average breakdown of wedding costs. According to their website, WeddingWire is “the largest and most trusted global marketplace connecting engaged couples with local wedding professionals”. WeddingWire published these data on a blog post dated December 2020: “We surveyed thousands of couples around the country in our WeddingWire Newlywed Report to share their wedding budgets with us”.

### Question: Is there one (or more) question addressed by the visualization?

The question is very clear: what is the budget breakdown per category of expenses for an average wedding?

### Data: Is the data quality appropriate?

Accuracy: data are comparable and the values are reasonable according to common judgment.

Completeness: data are complete, several categories are reported and we can assume that the list is exhaustive.

Consistency: the percentages of the categories correctly sum to 100%. The meaning of additional costs is unclear.

Currency: data are referred to the year 2020, so it is updated.

Credibility: the source is mentioned at the bottom and they are domain experts. We don't know how many couples answered the survey.

Understandability: data are understandable, but it is better to report absolute numbers instead of percentages.

Precision: a higher precision, maybe to the first decimal digit, would be more appropriate as many values are equal.

### Visual Proportionality: Are the values encoded in a uniformly proportional way?

Not at all, as the lengths of the bars representing 1% and 2% (or 2% and 3%) are almost equal.

### Visual Utility: All the elements in the graph convey useful information?

Several elements are useless: the colored background, the icons of the categories, the icon at the bottom-right, the textual comment, the rectangle around the title.

### Visual Clarity: Are the data in the graph clearly identifiable and understandable (properly described)?

The usage of direct labeling is appropriate and very clear. However, the meaning of the different colors associated with the bars is not clear.

## Design

Design the visualization based on the following data structure

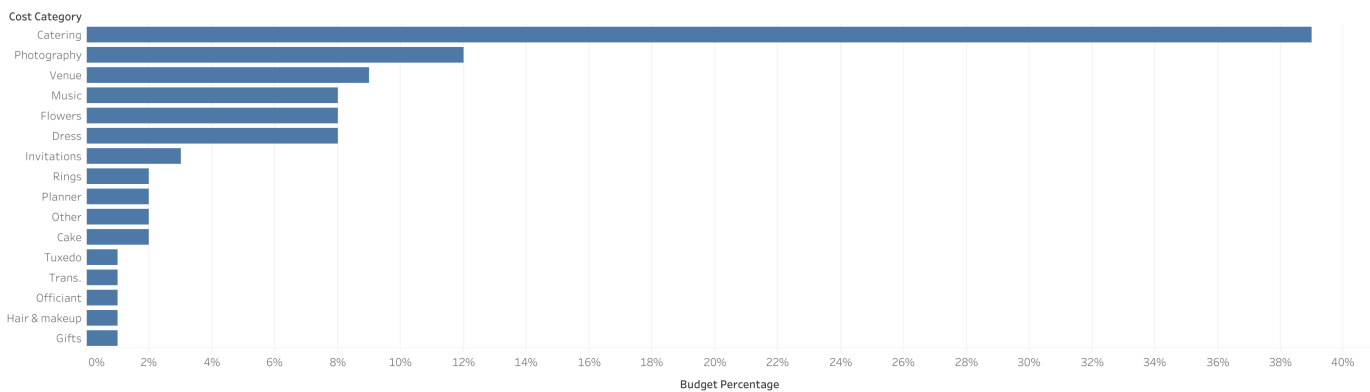
Field	Dim./Measure	Description
BUDGET_PERCENTAGE	Measure	Percentage of the total budget
COST_CATEGORY	Dimension	The categories of wedding expenses

### Design schema

Schema	Details
Columns:	SUM(BUDGET_PERCENTAGE)
Rows:	COST_CATEGORY
Graph type:	Bar
Color:	Default
Size:	Default
Label:	Default

### Sketch of the resulting graph

Bar chart

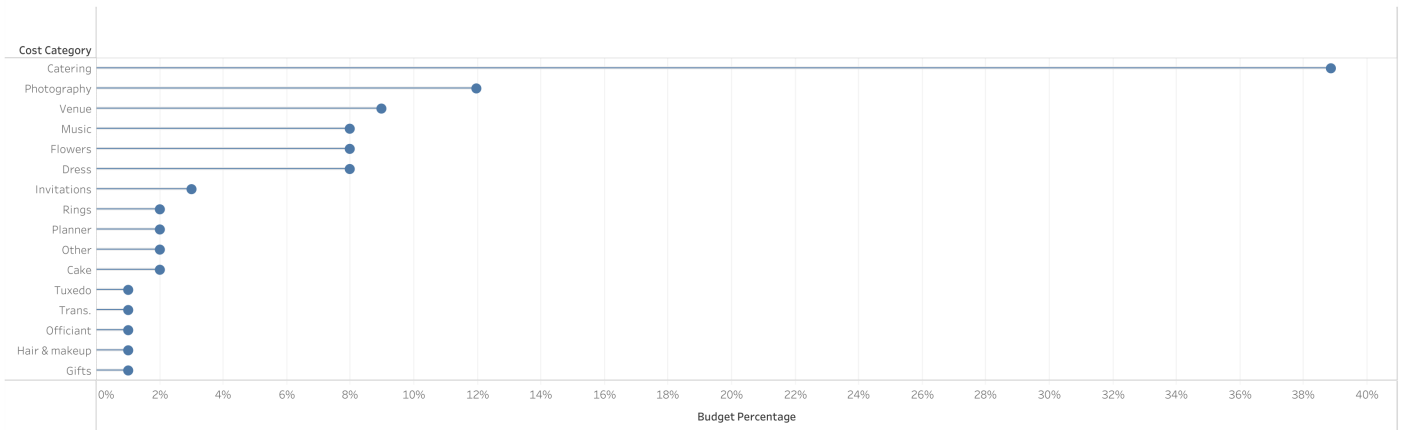


### Design schema

Schema	Details
Columns:	SUM(BUDGET_PERCENTAGE), SUM(BUDGET_PERCENTAGE)
Rows:	COST_CATEGORY
Graph type:	Bar, Circle
Color:	Default
Size:	Smaller, Default
Label:	Default

## Sketch of the resulting graph

Lollipop

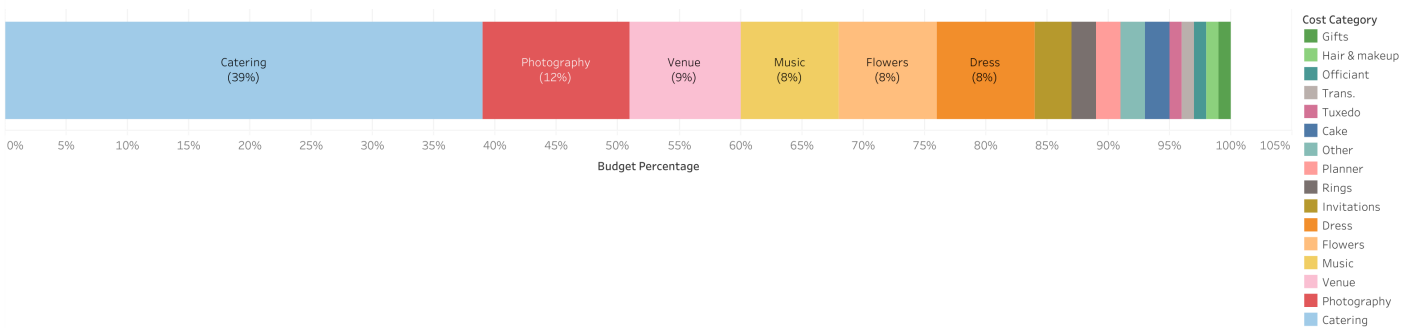


## Design schema

Schema	Details
Columns:	SUM(BUDGET_PERCENTAGE)
Rows:	-
Graph type:	Bar
Color:	COST_CATEGORY
Size:	Default
Label:	COST_CATEGORY, SUM(BUDGET_PERCENTAGE)

## Sketch of the resulting graph

Stacked bars

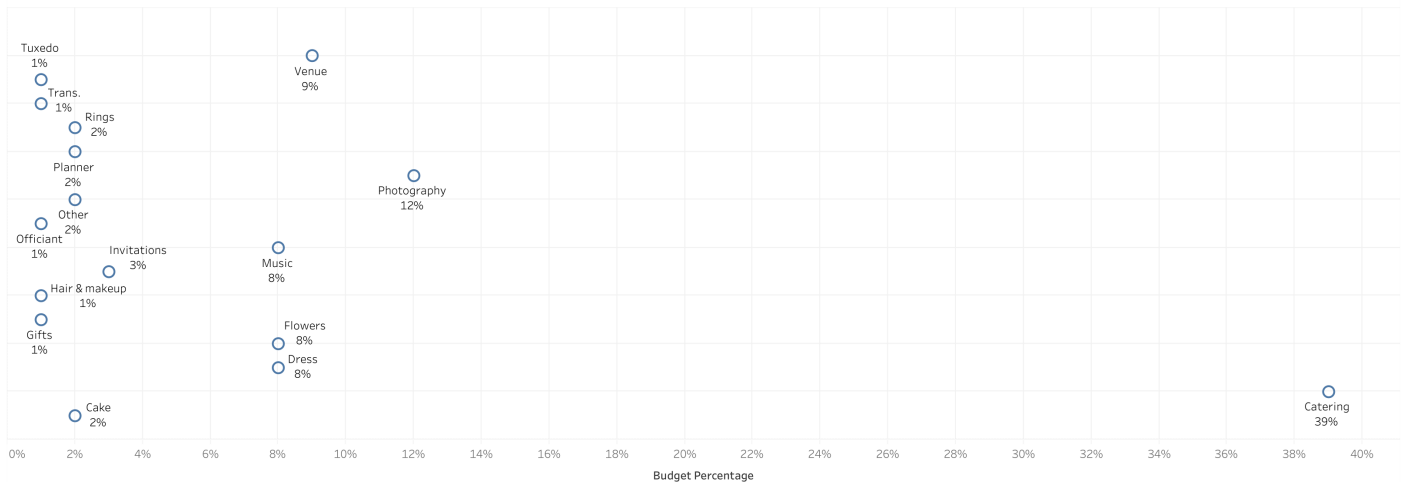


## Design schema

Schema	Details
Columns:	SUM(BUDGET_PERCENTAGE)
Rows:	INDEX(COST_CATEGORY)
Graph type:	Shape
Color:	Default
Size:	Bigger
Label:	COST_CATEGORY, SUM(BUDGET_PERCENTAGE)

## Sketch of the resulting graph

Dots with jitter



## Theory

Which one of the following examples is **NOT** related to a Gestalt principle?

- the points of a group are enclosed by a fine line
- the color of the legend is similar to the color of the elements of the graph
- the direct labeling technique improves the readability of the visualization
- *the bars representing smaller values are shorter*
- the points of a data series are connected