

⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > EDA: Examining Relationships > Case C→Q > Case C→Q: Introduction

🔖 Bookmark this page

Case C→Q: Introduction

Learning Objective: Compare and contrast distributions (of quantitative data) from two or more groups, and produce a brief summary, interpreting your findings in context.

Case C→Q: Categorical Explanatory Variable and Quantitative Response Variable

Recall the role-type classification table for framing our discussion about the relationship between two variables:

		Response	
		Categorical	Quantitative
Explanatory	Categorical	C→C	C→Q
	Quantitative	Q→C	Q→Q

We are now ready to start with Case C→Q, exploring the relationship between two variables where the explanatory variable is categorical, and the response variable is quantitative. As you'll discover, exploring relationships of this type is something we've already discussed in this course, but we didn't frame the discussion this way.

Example: Hot Dogs

Background: People who are concerned about their health may prefer hot dogs that are low in calories. A study was conducted by a concerned health group in which 54 major hot dog brands were examined, and their calorie contents recorded. In addition, each brand was classified by type: beef, poultry, and meat (mostly pork and beef, but up to 15% poultry meat). The purpose of the study was to examine whether the **number of calories** a hot dog has is related to (or affected by) its **type**. (Reference: Moore, David S., and George P. McCabe (1989). *Introduction to the Practice of Statistics*. Original source: *Consumer Reports*, June 1986, pp. 366-367.)

Answering this question requires us to examine the relationship between the categorical variable, Type and the quantitative variable Calories. Because the question of interest is whether the type of hot dog affects calorie content,

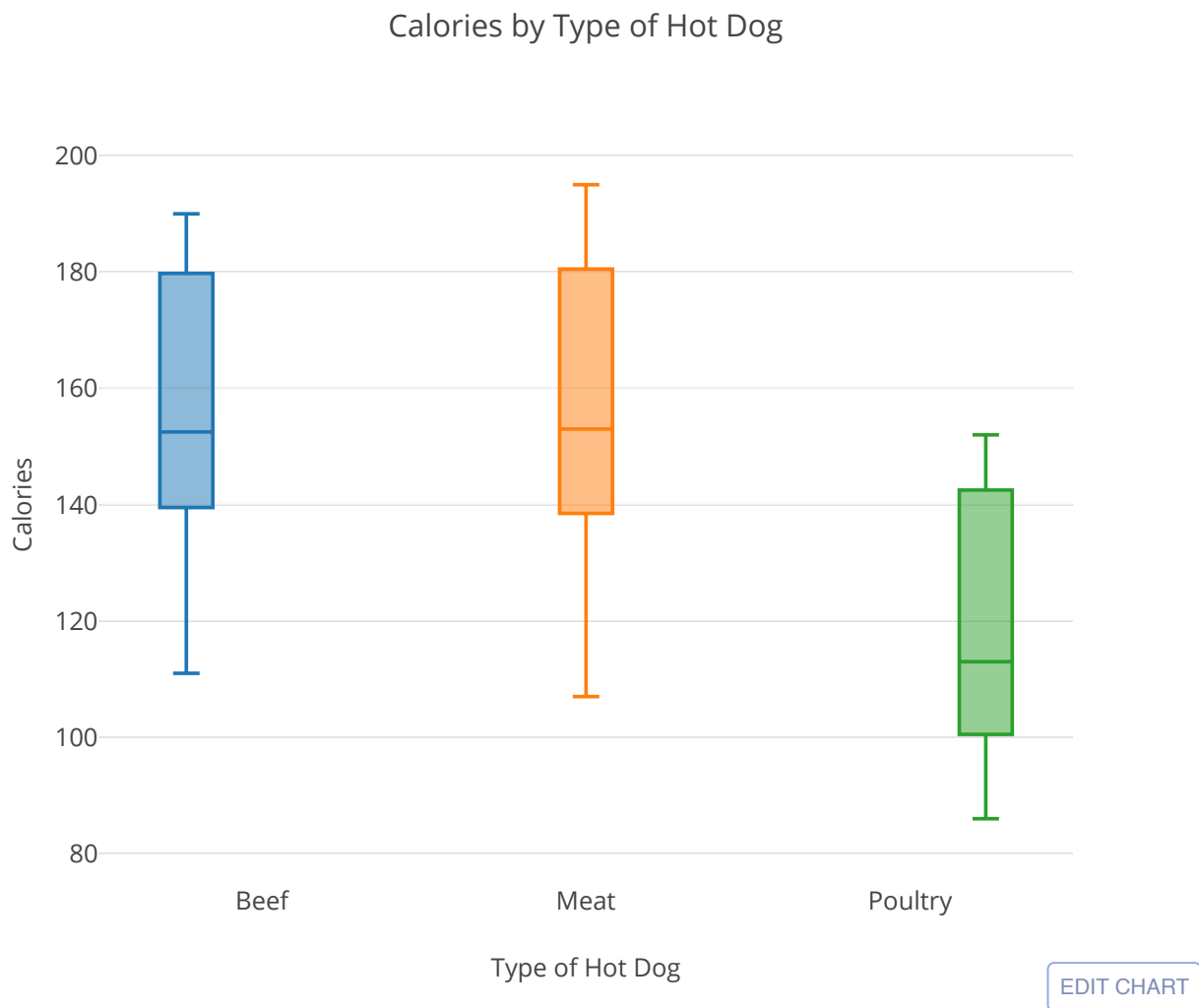
- the **explanatory** variable is **Type**, and
- the **response** variable is **Calories**.

Here is what the raw data look like:

	Type	Calories
Brand 1	Beef	186
Brand 2	Poultry	129
Brand 3	Beef	181
Brand 4	Meat	173
.	.	.
.	.	.
.	.	.
Brand 54	Poultry	144

The raw data are a list of types and calorie contents, and are not very useful in that form. To explore how the number of calories is related to the type of hot dog, we need an informative visual display of the data that will compare the three types of hot dogs with respect to their calorie content.

The visual display that we'll use is **side-by-side boxplots** (which we've seen before). The side-by-side boxplots will allow us to **compare the distribution** of calorie counts within each category of the explanatory variable, hot dog type:



As before, we supplement the side-by-side boxplots with the descriptive statistics of the calorie content (response) for each type of hot dog separately (i.e., for each level of the explanatory variable separately):

Statistic	Beef	Meat	Poultry
min	111	107	86
Q1	139.5	138.5	100.5
Median	152.5	153	113
Q3	179.75	180.5	142.5
Max	190	195	152

Let's summarize the results we got and interpret them in the context of the question we posed:

By examining the three side-by-side boxplots and the numerical summaries, we see at once that poultry hot dogs, as a group, contain fewer calories than those made of beef or meat. The median number of calories in poultry hot dogs (113) is less than the median (and even the first quartile) of either of the other two distributions (medians 152.5 and 153). The spread of the three distributions is about the same, if IQR is considered (all slightly above 40), but the (full) ranges vary slightly more (beef: 80, meat: 88, poultry: 66). The general recommendation to the health-conscious consumer is to eat poultry hot dogs. It should be noted, though, that since each of the three types of hot dogs shows quite a large spread among brands, simply buying a poultry hot dog does not guarantee a low-calorie food.

What we learn from this example is that when exploring the relationship between a categorical explanatory variable and a quantitative response (Case C→Q), we essentially **compare the distributions of the quantitative response for each category of the explanatory variable** using side-by-side boxplots supplemented by descriptive statistics. Recall that we have actually done this before when we talked about the boxplot and argued that boxplots are most useful when presented side by side for comparing distributions of two or more groups. This is exactly what we are doing here!

Open Learning Initiative [🔗](#)



[🔗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [🔗](#).

© All Rights Reserved