DMC 2015

So You Want To Be A Data Miner?

ISU Team

Presentation by Ian Mouzon

April and May, 2015

The Data Mining Cup

What is it?

The Data Mining Cup is a yearly competition hosted by prudsys (all lower case) a German analytics company focused on marketplace behavior.

> The DATA MINING CUP (DMC for short) has been inspiring students around the world to pursue intelligent data analysis since the year 2000. In 2014 over one thousand students from about 100 universities in 28 countries took part in the competition. The best teams will be invited to Berlin for the awards ceremony at the prudsys personalization summit.

The Data Mining Cup

What is ISU's history in the DMC

In 2013 we had six team members (all from the STAT department and most associated in some way with Dr. Vardeman and STAT 602).


Wen Zhou


Cory Lanker


Fangfang Liu


Jia Liu


Ian Mouzon


Wei Zhang

Our team came in fifth place and some of our members were able to travel to Berlin for the awards ceremony.

The Data Mining Cup

What is ISU's history in the DMC



The Data Mining Cup

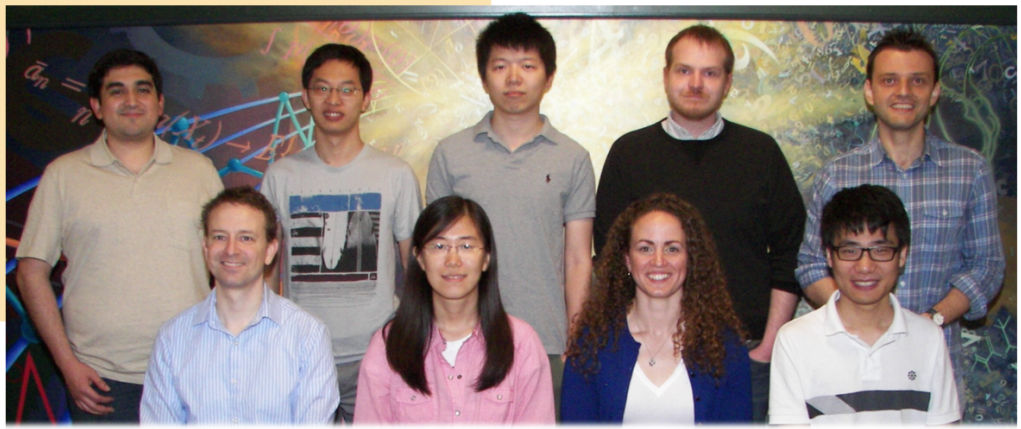
What is ISU's history in the DMC

In 2014, the team we put together was able to win first place

+ DMC 2014 Solution

Team Uni_Iowa_State_1

Iowa State University
Department of Statistics



Presenters: Guillermo Basulto, Fan Cao, Cory Lanker, Xin Yin

Absent: Zoe Cheng, Marius Dragomiroiu, Jessica Hicks, Ian Mouzon, Lanfeng Pan

Predicting Online Orders

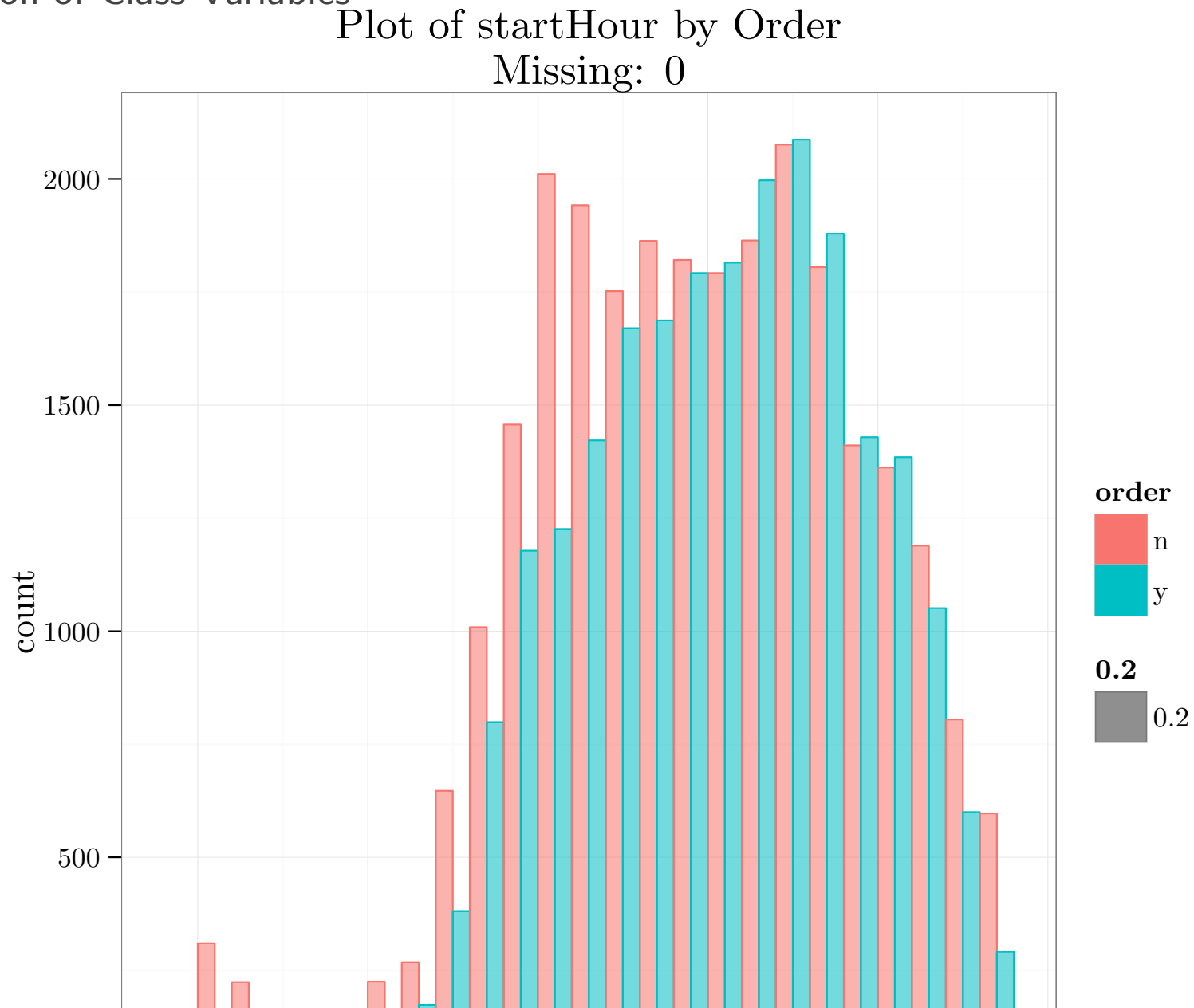
The Data Mining Cup, Sponsored by prudsys

A Data Mining Competition with two tasks:

1. Use 50,000 customers session records to predict the ordering status of new records,
2. Be able to determine in real-time whether or not a session will end with the customer placing an order.

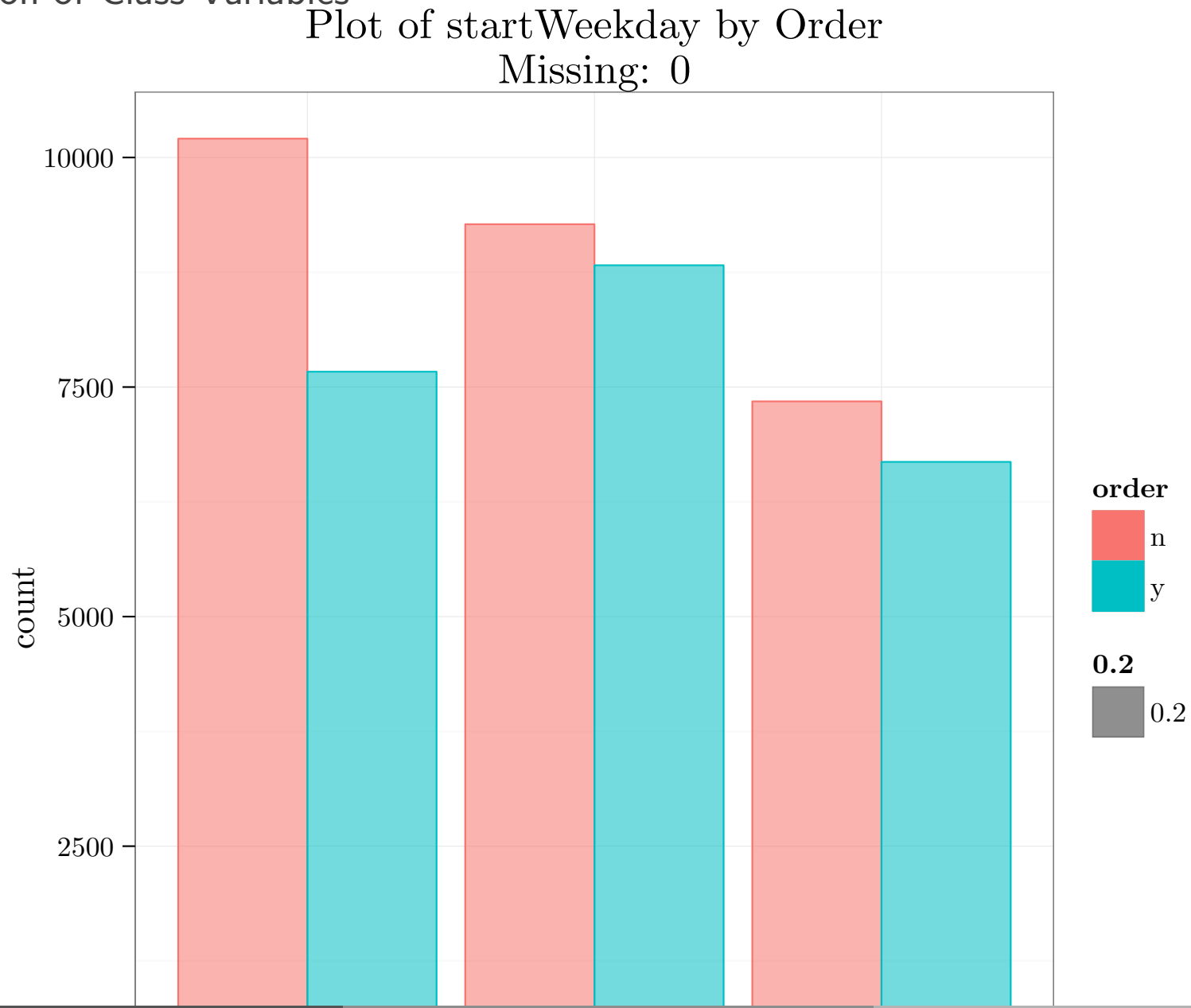
Overview of Dataset

Description of Class Variables



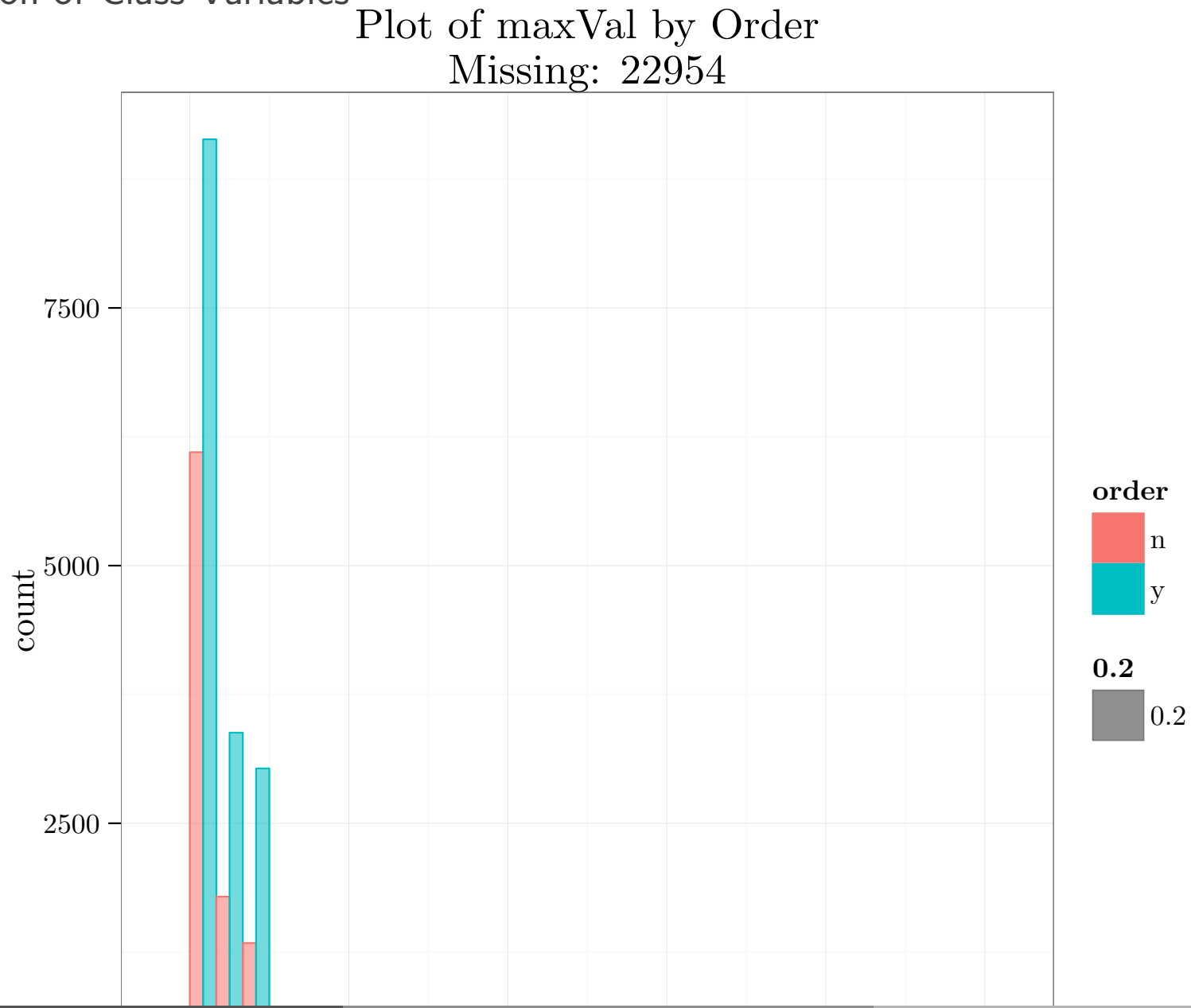
Overview of Dataset

Description of Class Variables



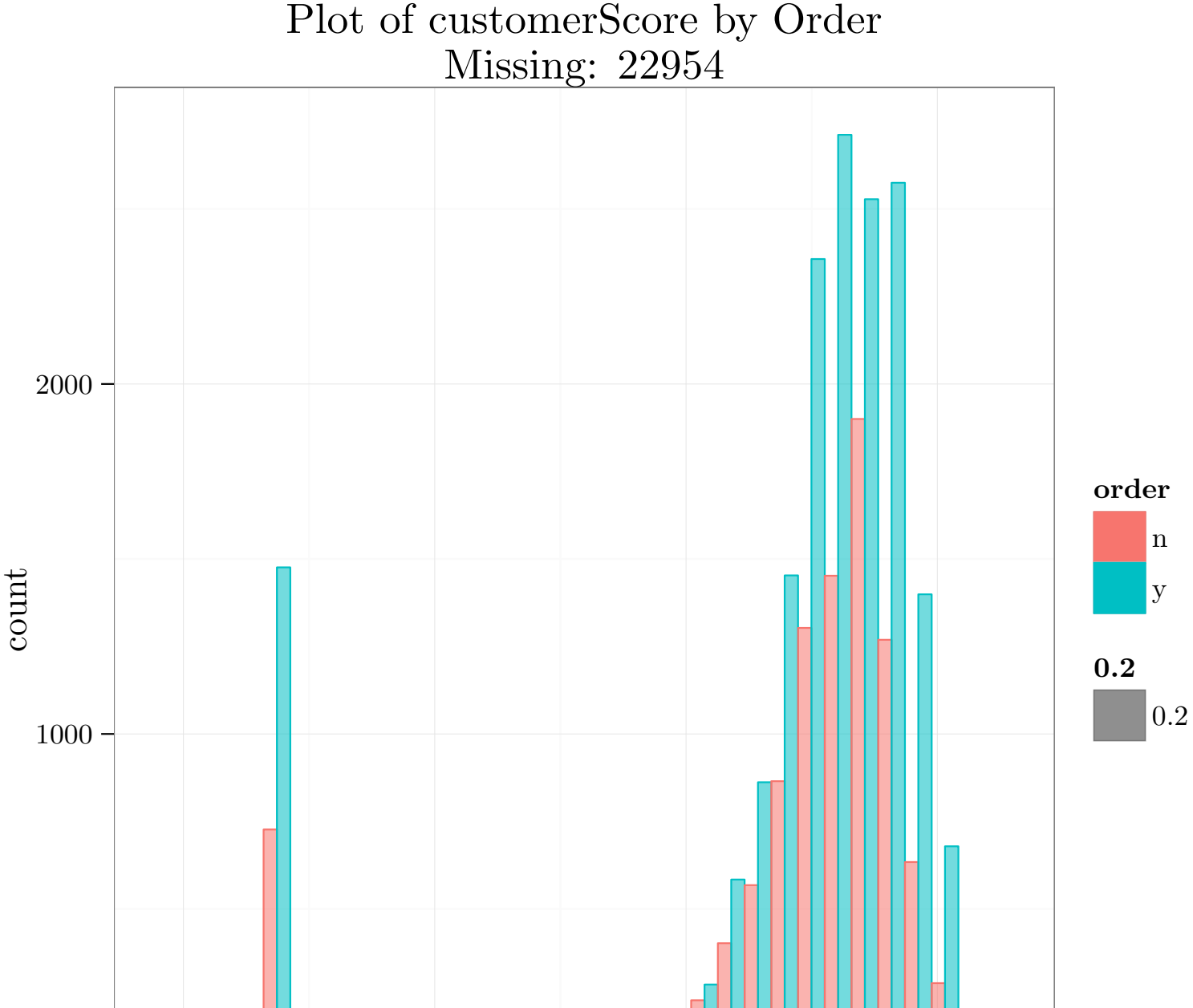
Overview of Dataset

Description of Class Variables



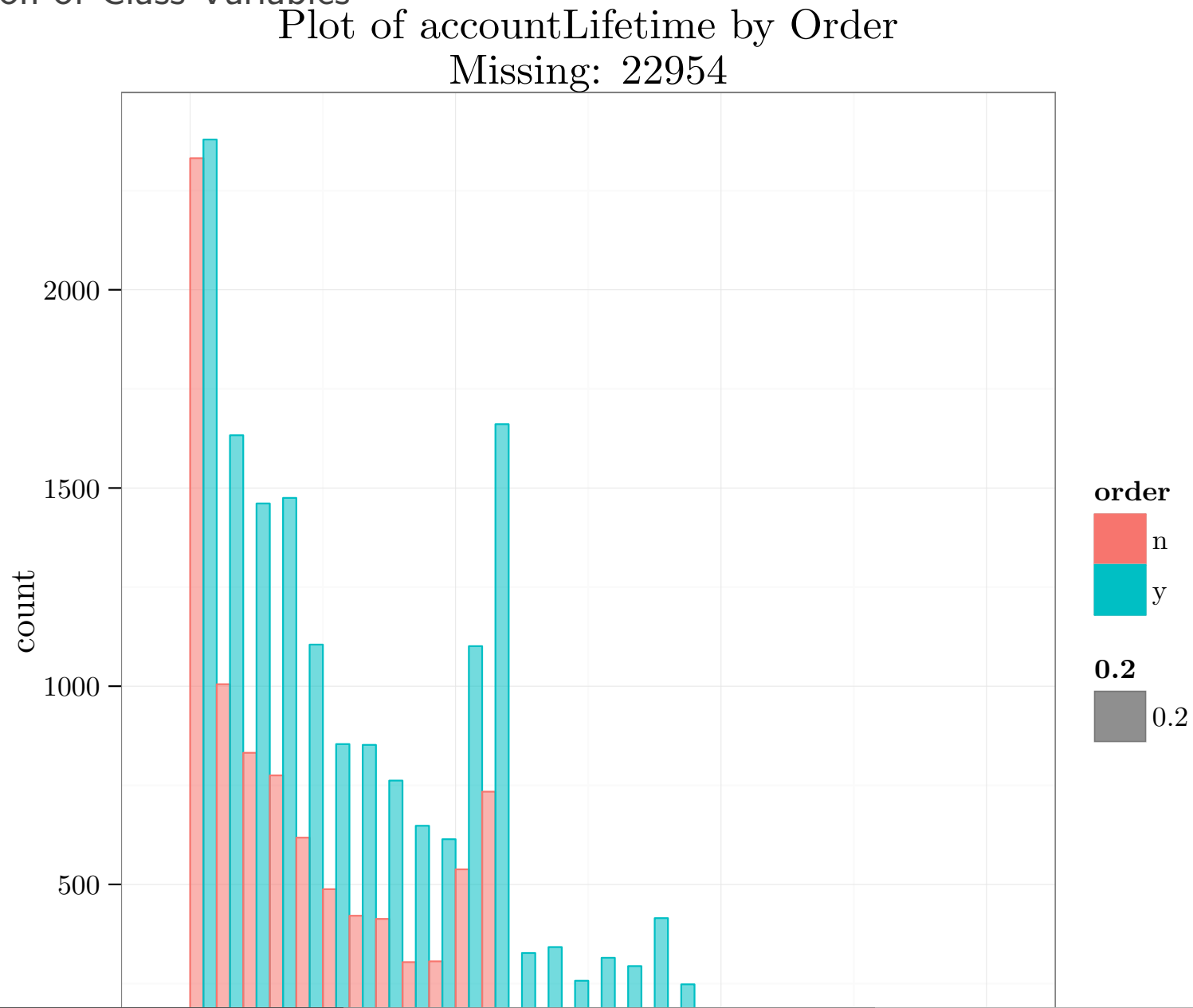
Overview of Dataset

Description of Class Variables



Overview of Dataset

Description of Class Variables

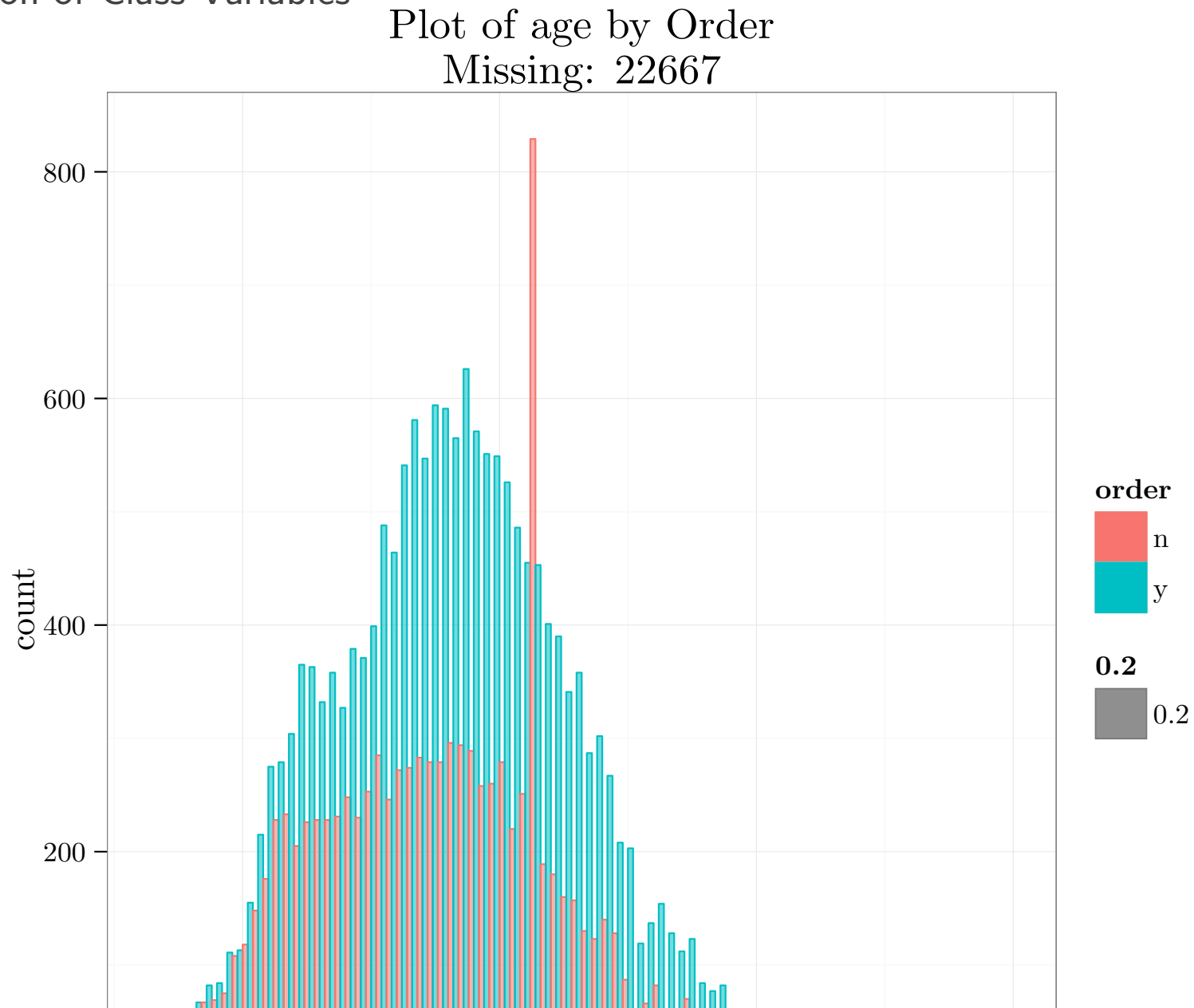


Overview of Dataset

Description of Class Variables

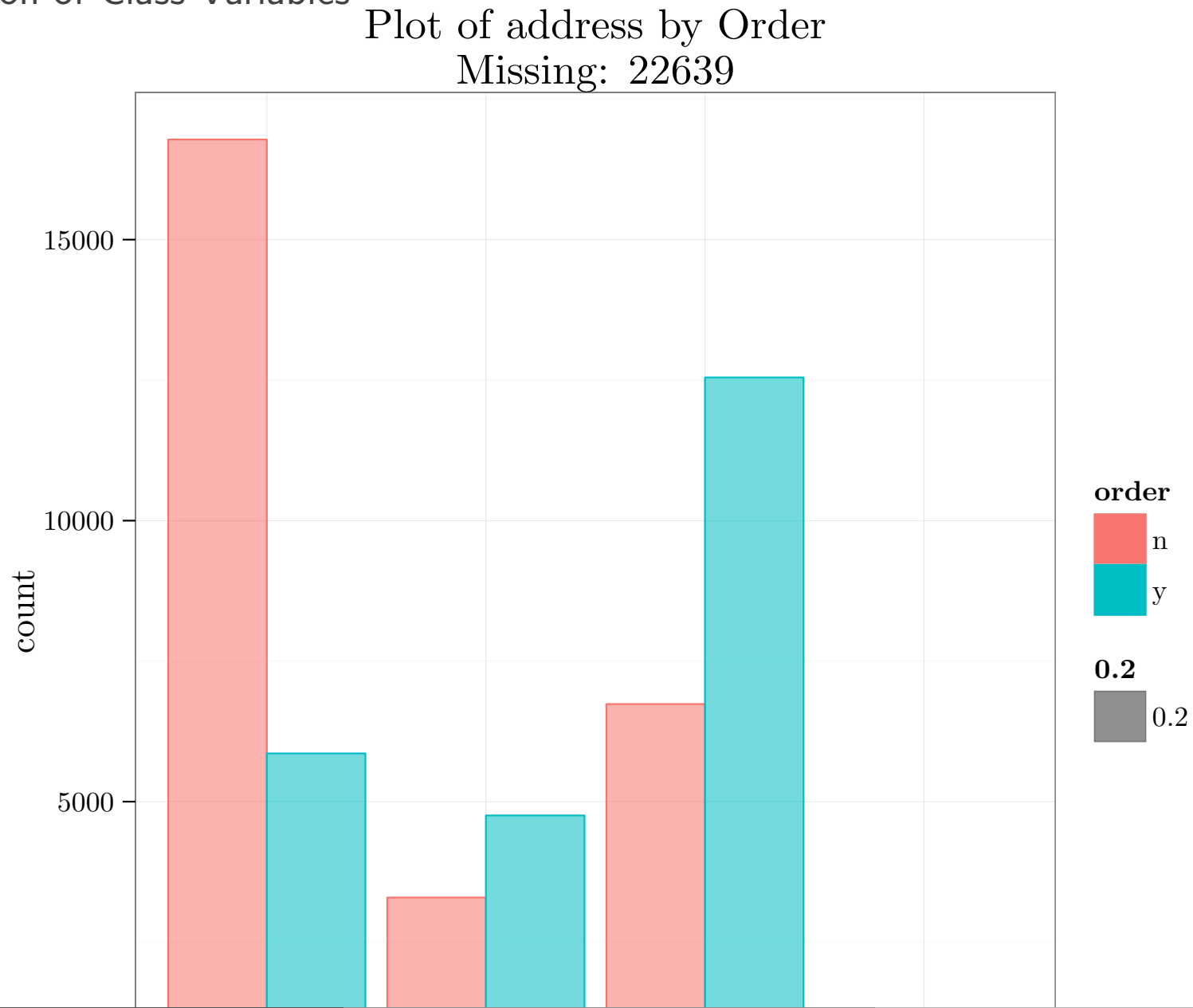
Overview of Dataset

Description of Class Variables



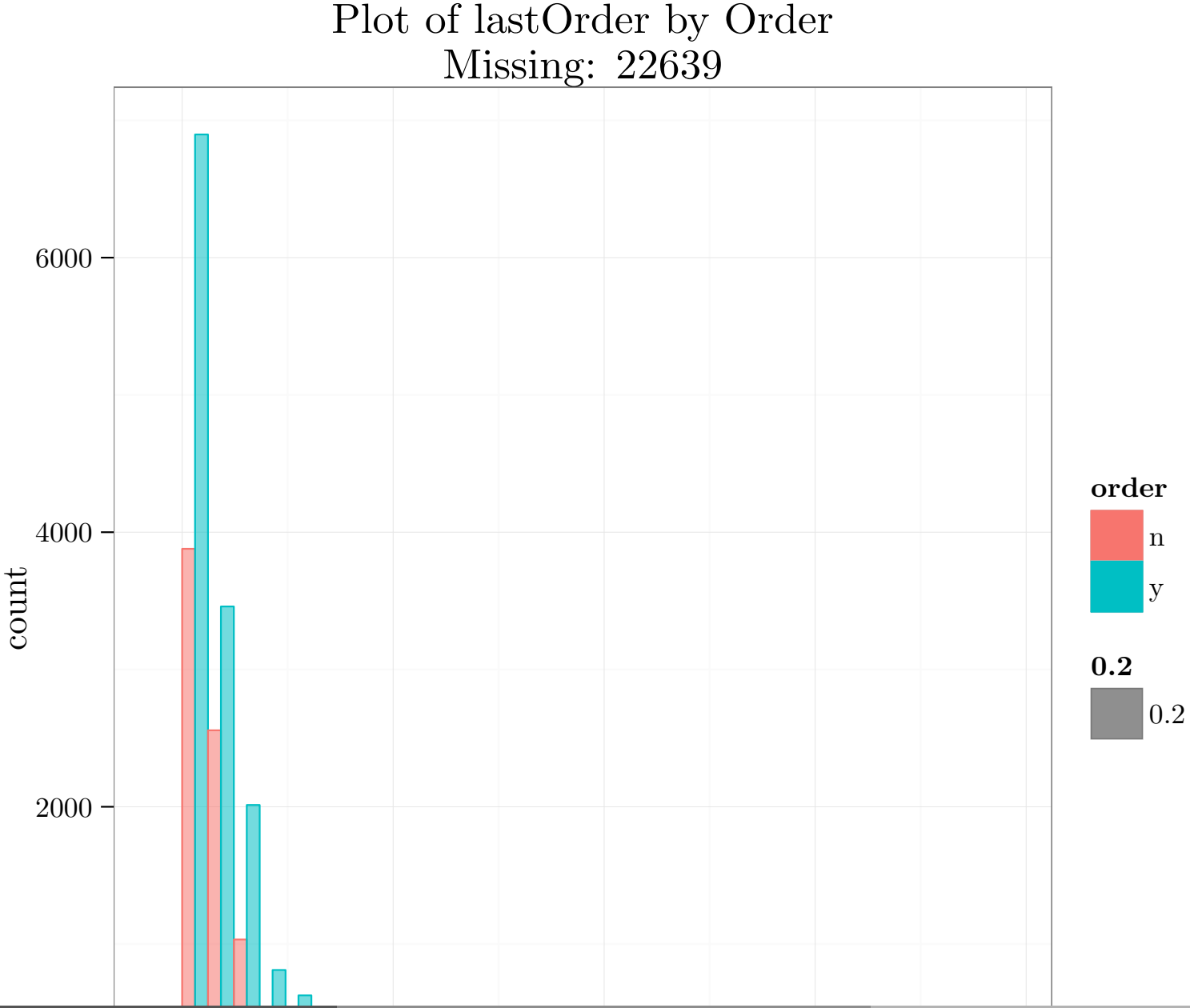
Overview of Dataset

Description of Class Variables



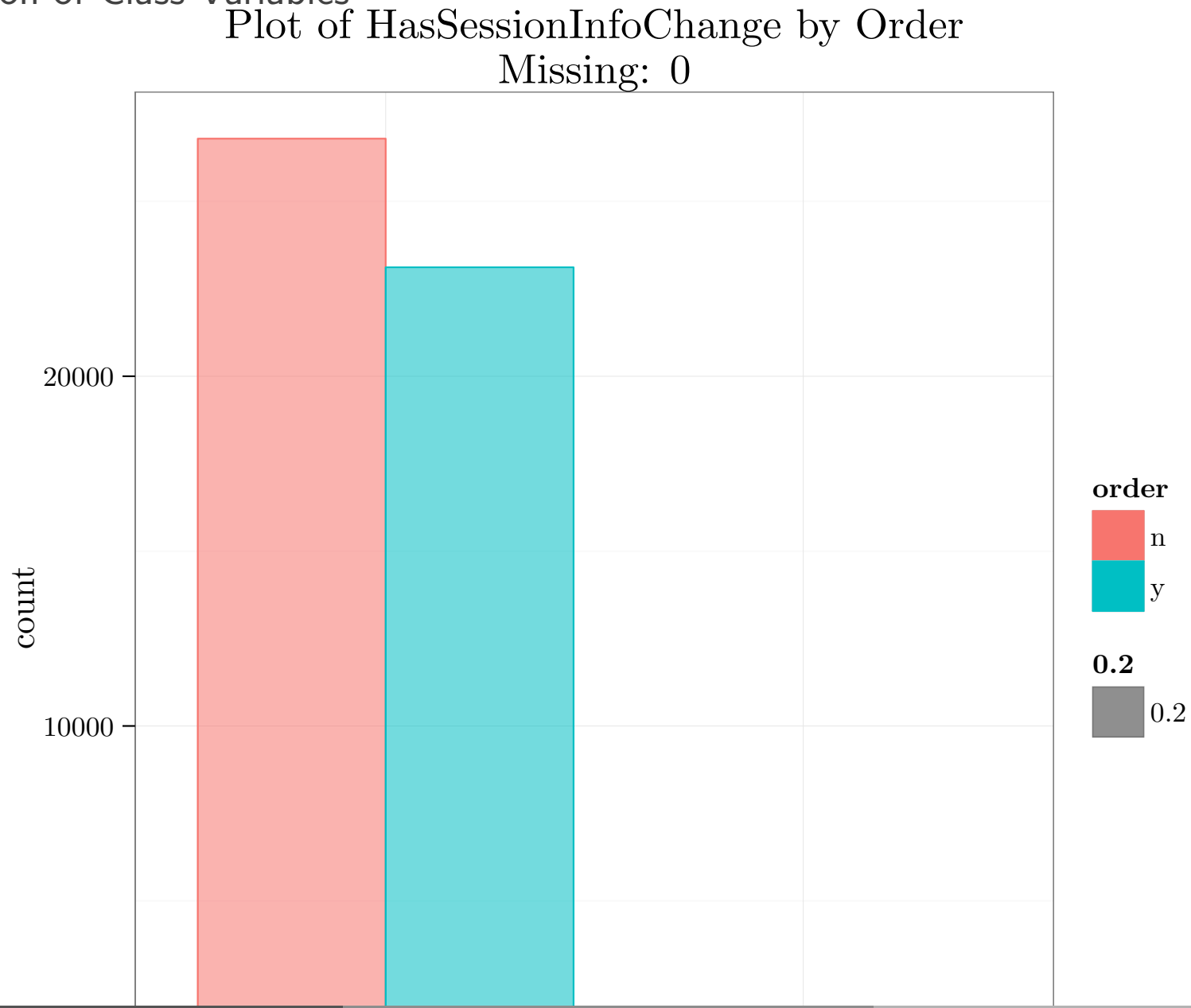
Overview of Dataset

Description of Class Variables



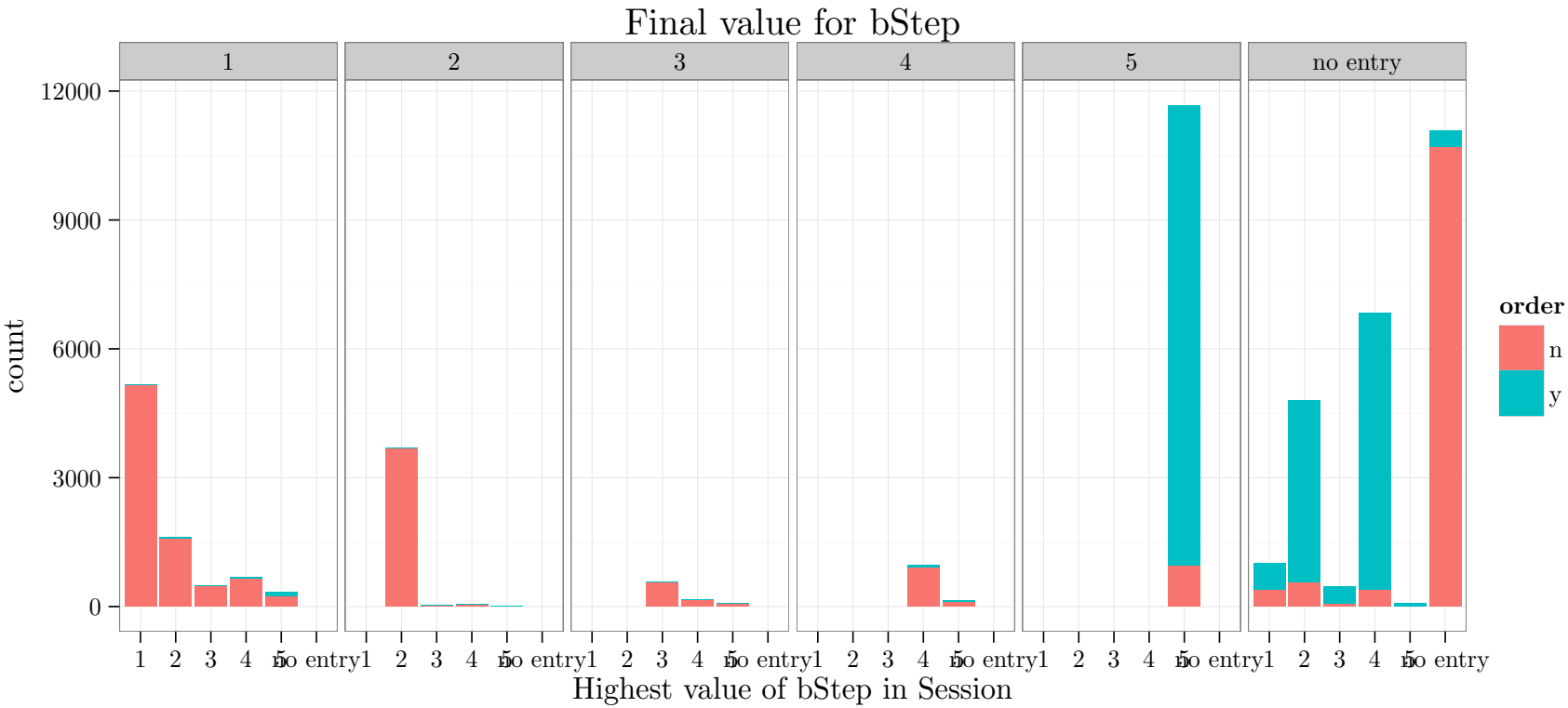
Overview of Dataset

Description of Class Variables



Overview of the Dataset

Key Feature: bStep



- This has split the sessions into two groups: A haystack with a few needles and a needle stack with a little hay.
- We can sort out the details using the rest of the data.