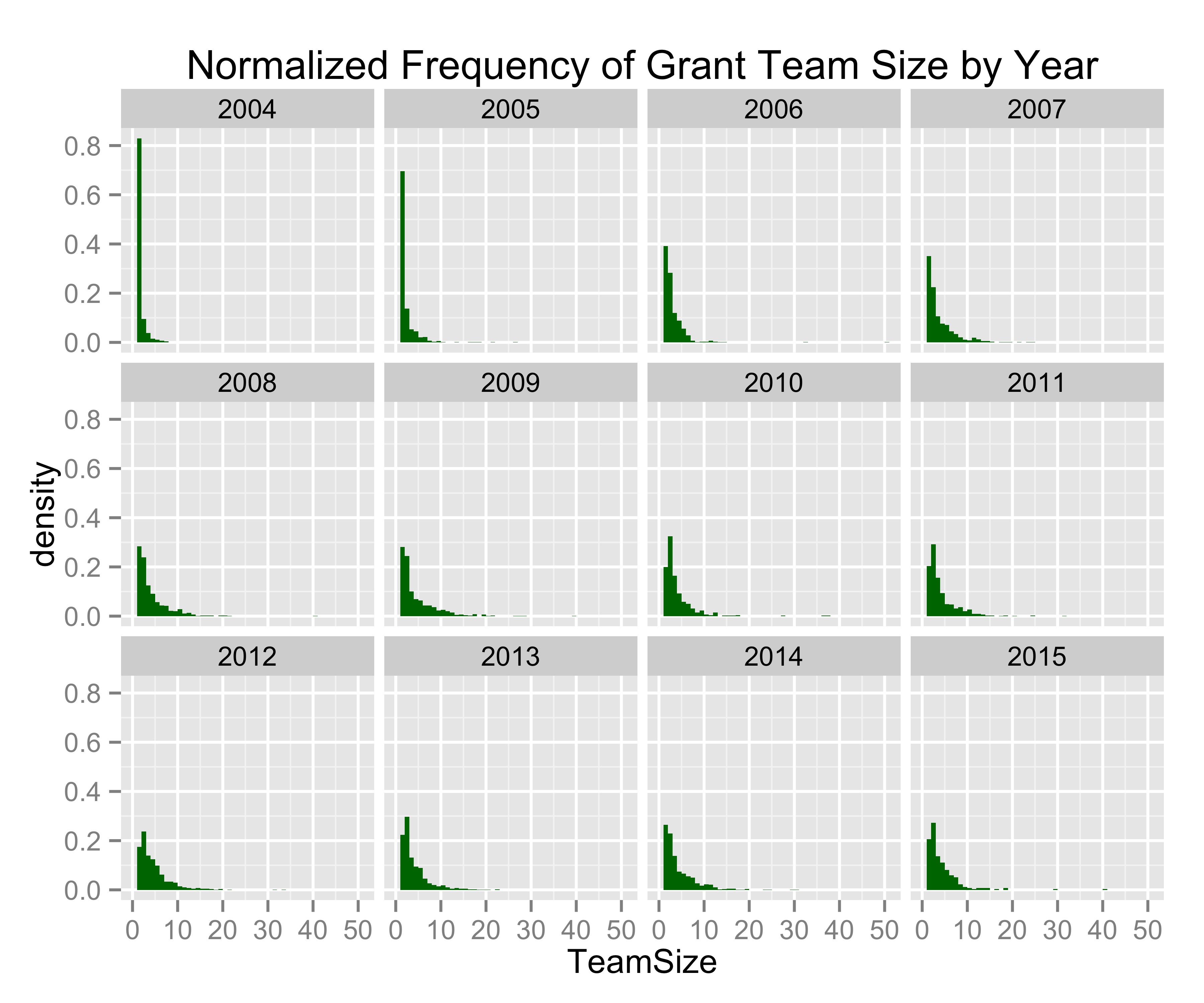
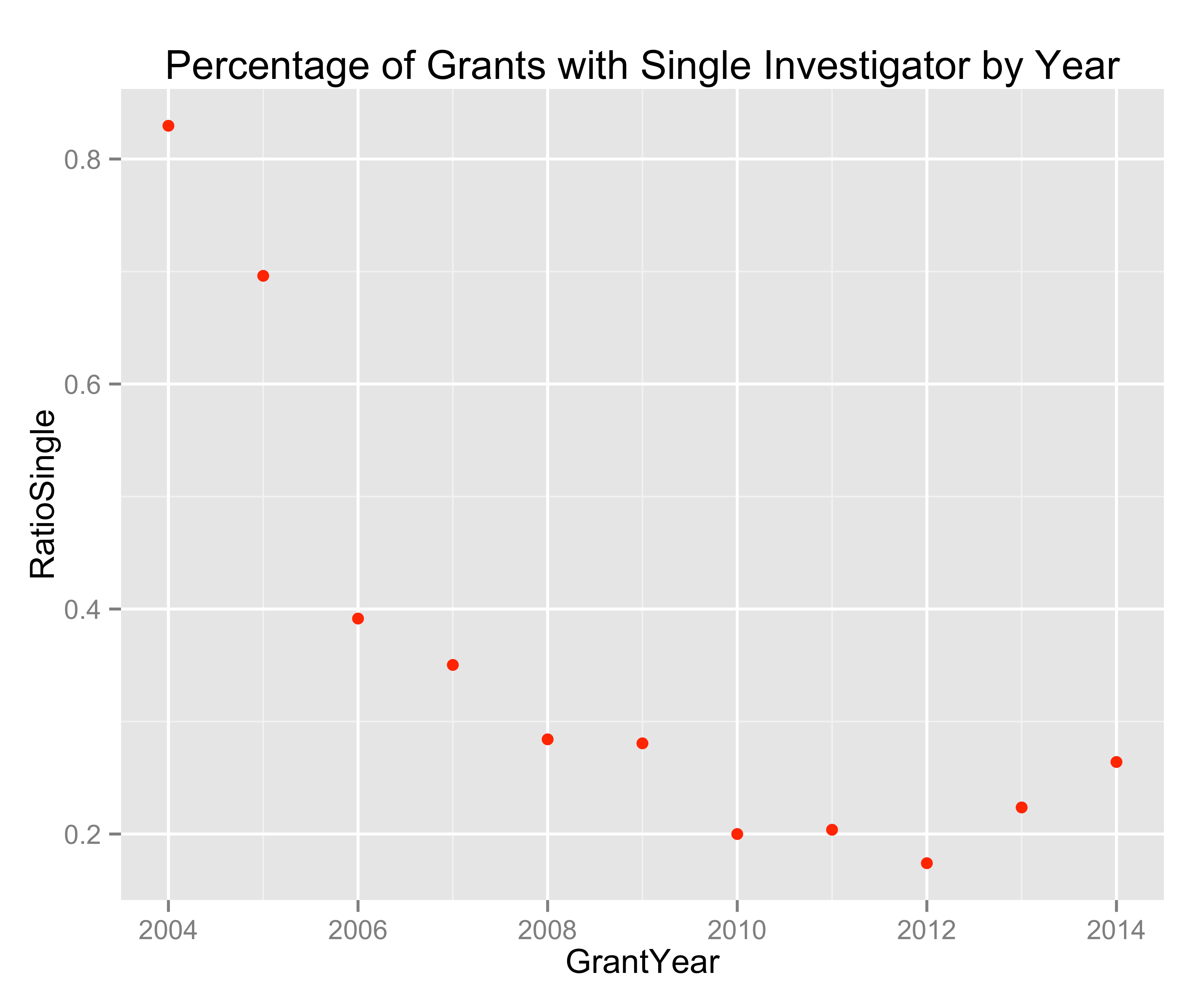
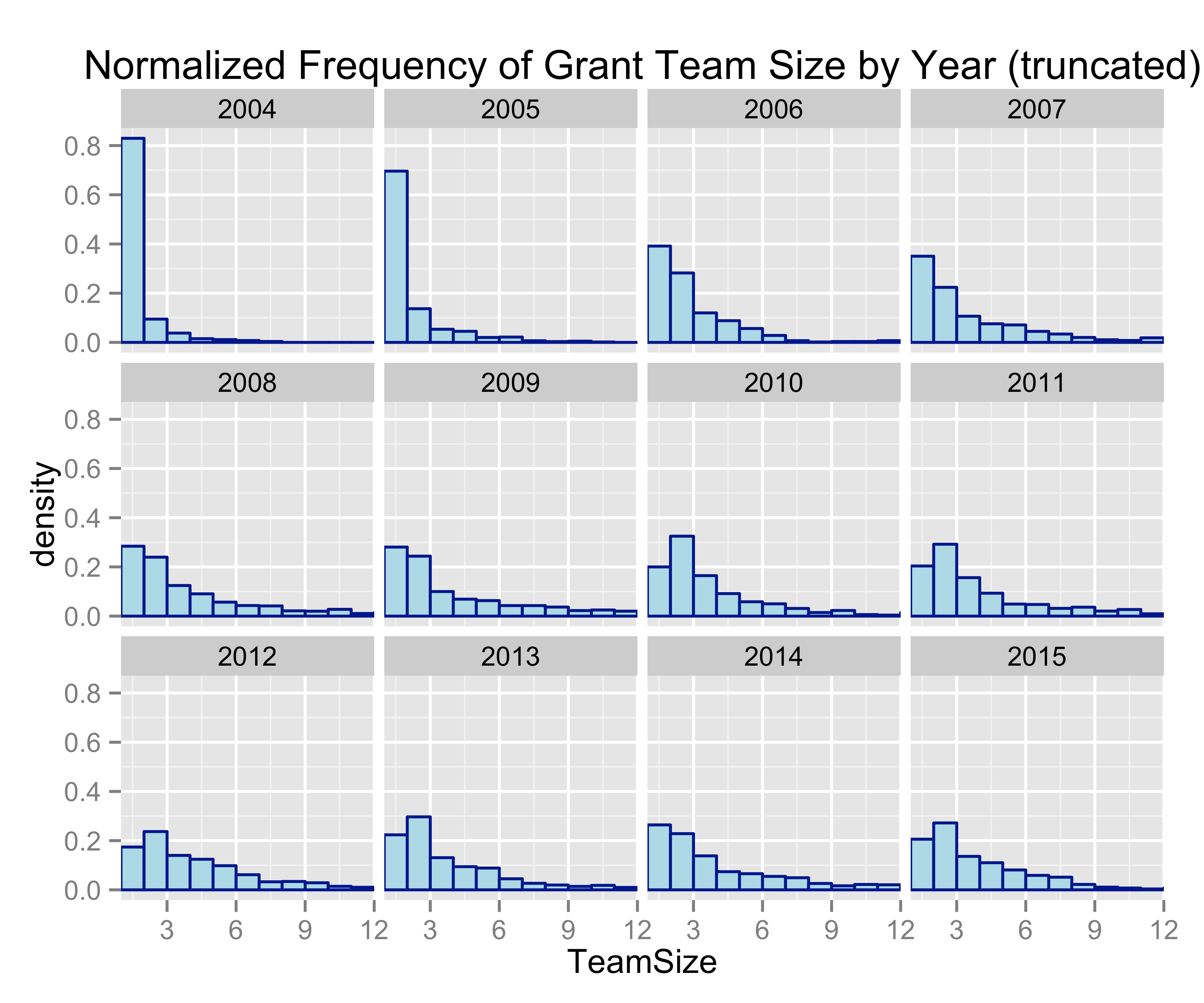
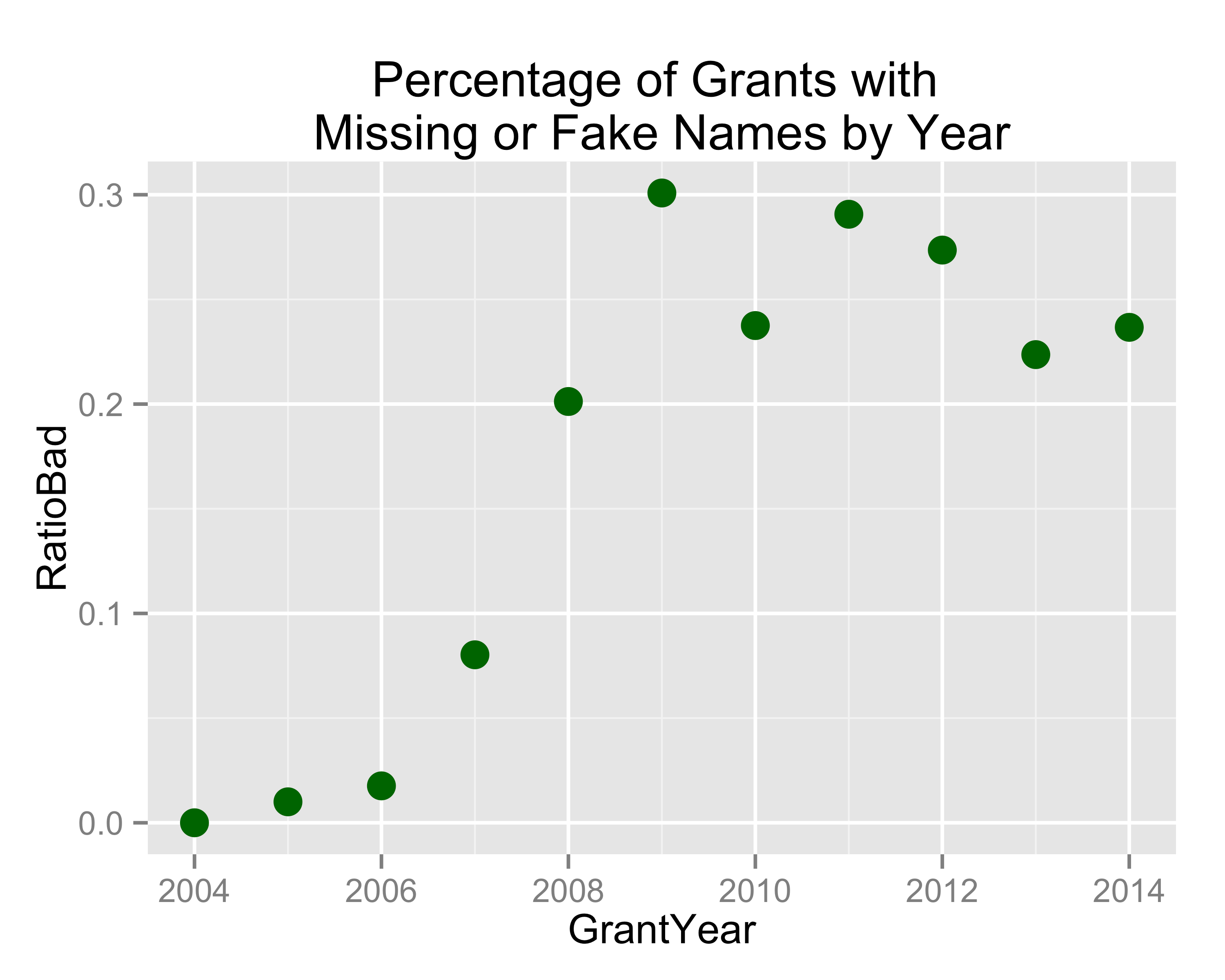
Preliminary Data Analysis

We recorded the number of investigators participating in each grant (TeamSize). Then for each year, we collected the frequencies of each team size and normalized them in the above density histogram. This means that the x-values represent the team size, and y-values represent the percent of grants from that year that had x participants. The most interesting temporal changes happen at lower team sizes, so to better visualize these changes we look at another table showing only team sizes 1-12.

Between these two graphs, we can see that in 2004 and 2005, the majority of grants were proposed with only one investigator. This changes in 2006, when the percentage of grants with only one investigator drops below 40%.

The year 2010 marks the first year that there are more grants with two investigators than grants with only one investigator. Note that between 2009 and 2010 VCU received the CCTR grant.

On possible explanation for this drop in individual investigators is that perhaps there was a change in how grant proposals were filed. Maybe in 2004 and 2005 it was not required to list all of the members of the team, and in later years there was a policy change.

Notice in the graph to the right that there is an increasing trend to enter placeholder names instead of listing the actual names of the investigators involved in the grant. These grants with placeholders probably account for the enormous differences we see between the earlier years and the later years. From 2008 onward, between 20% and 30% of the grants proposed use at least one placeholder name.

This makes data analysis difficult if we want to get a better understanding of the relationship between collaboration and grant success. Ideally of course every grant proposal would be submitted with the identification numbers (V-numbers in our case) of every investigator who collaborated on a project.

We need to decide how to treat this fault in the data. Maybe it is best to ignore years 2004-2006.

Of grants that **were awarded**, **19.03%** had at least one name holder in the team. Of grants that **were not awarded**, **19.13%** had at least one name holder in the name. There does not seem to be a correlation between missing names and the success of the grant proposal, so we have decided to eliminate missing names and placeholder names from the investigator lists. The consequence of this is that from our network analyses we also remove any edges involving an investigator with a missing name or a name holder.