Five defenses before Thanksgiving:

* Dan – 500-year-old book. Created dataset corpus with shiny app to see how keywords interact
* Daniel – loss rates of loan default
  + Embargoed data - proprietary method, sensitive data
* Jerry – baseball data to look at performance and pay metrics based upon nationality
* Michael – load for power plants operating at capacity.
  + Embargoed data – location
* Mark – generic drugs and how they interact with patent lifespan

Deadline Rules:

1 week before semester end – final paper with all corrections, in proper format and proper font

3 weeks before semester end – presentation

6 weeks before semester end – submit final draft

Typically 6 to 7 drafts before final draft. IMRAD format.   
<https://docs.google.com/document/d/1aGxhfa9QZs0cdUwoDK5035GgxF_8A5kX7U9UQbkOOeQ/edit>

Introduction has varied between 3 and 32 pages (both were on the extreme side).

Method tends to be around 5 pages of text (where did data come from, what was used)

Discussion typically has a couple of graphs to focus on. Discuss conclusion. Care more about discussion than results.

References

Appendices

Method first, find data.   
Data first, find method.

620 as jumping off point for 630.

3-4 people on a committee. A chair/advisor, a DS professor as first reader, and a 3rd/4th person outside dept.

Split dataset by when it’s under a flag or certain condition. Call from two or three separate dataframes.