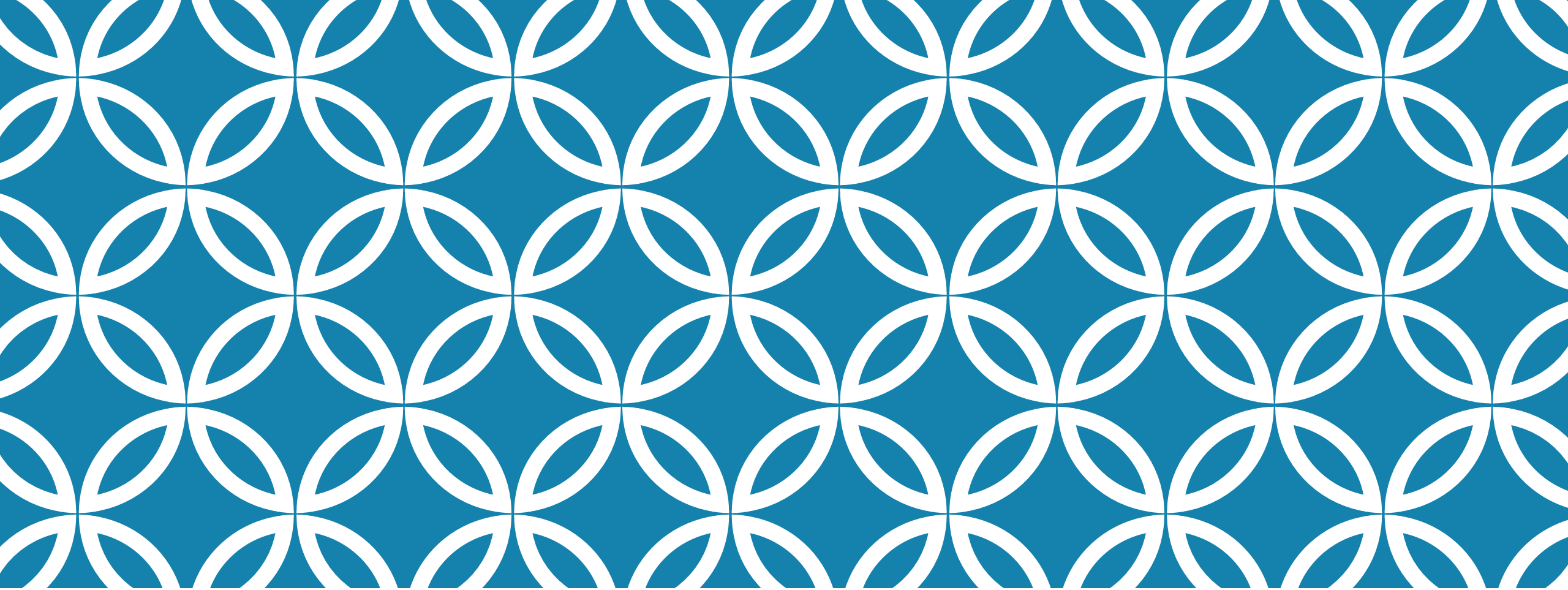


QUICK ANALYSIS OF NFL DRAFT DATA USING ARULES & GGPLOT2 PACKAGE IN R

Saad Yousef
(Twiter: SaadDataSci, GitHub: snyousef)
Job Seeker in Predictive Analytics, Data Analytics, Data Science

DESCRIPTION

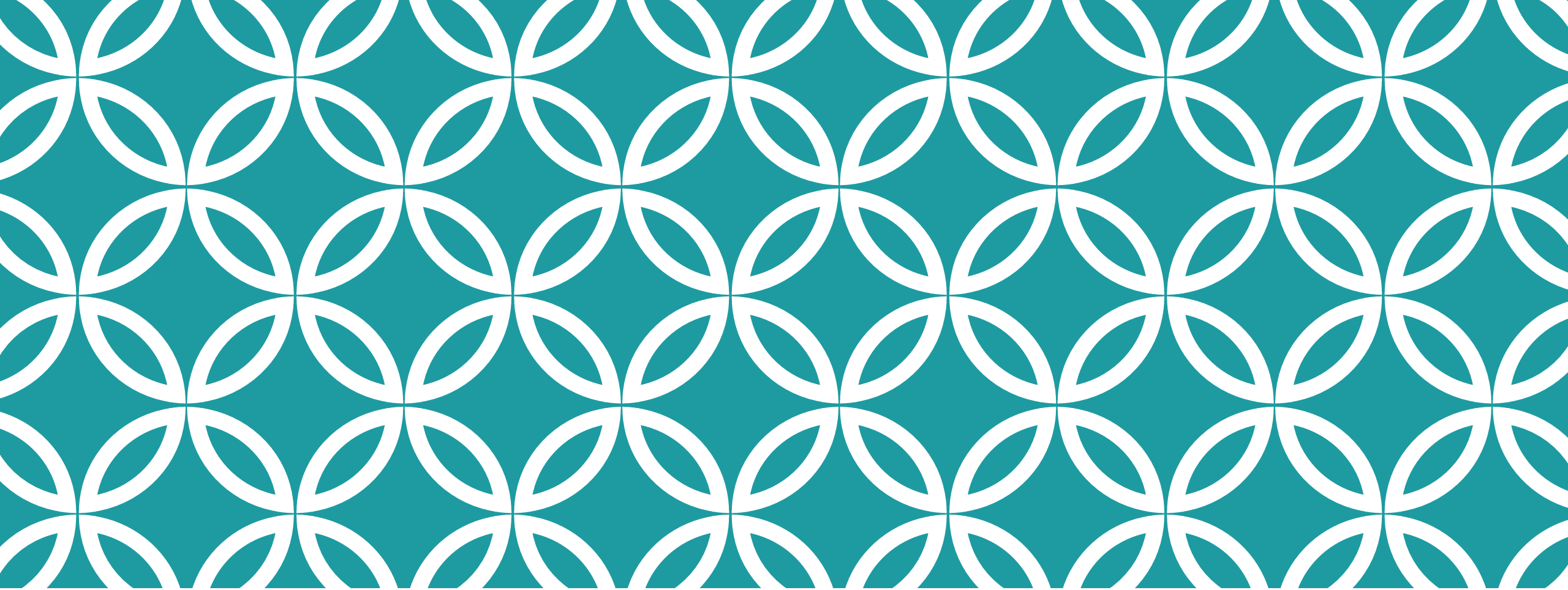
Took all draft data from years 2014-2018 and pared data down to four attributes:

- NFL Team (Arizona Cardinals, Dallas Cowboys, etc.)
- NCAAF Team (Alabama, Texas, etc.)
- NCAAF Team's Conference (ACC, Big 12, etc.)
- Player Position (WR, QB, etc.)

OBJECTIVE

Look for patterns in team's drafting habits. For example:

- Do certain NFL teams draft more players from a school than any other school?
- Do certain NFL teams draft more players from a conference than any other school?
- Do certain NFL teams draft more players at one position than any other team?
- Do players of a certain position come from one conference more than another?



EXPLORATORY DATA ANALYSIS

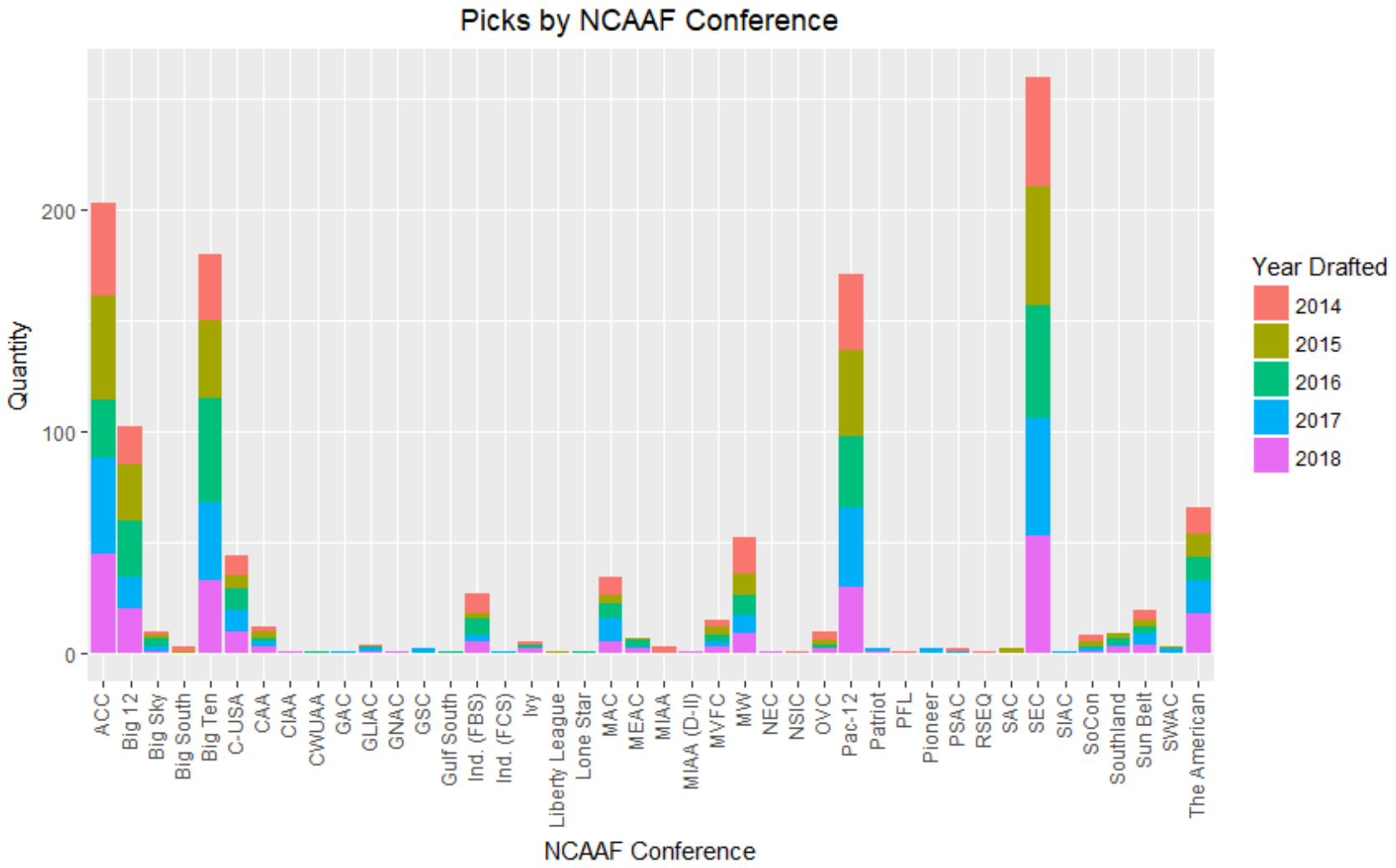
INITIAL OBSERVATIONS

SEC (surprise) is conference with most people drafted (then came ACC, Big Ten, Pac-12, Big 12 in that order)

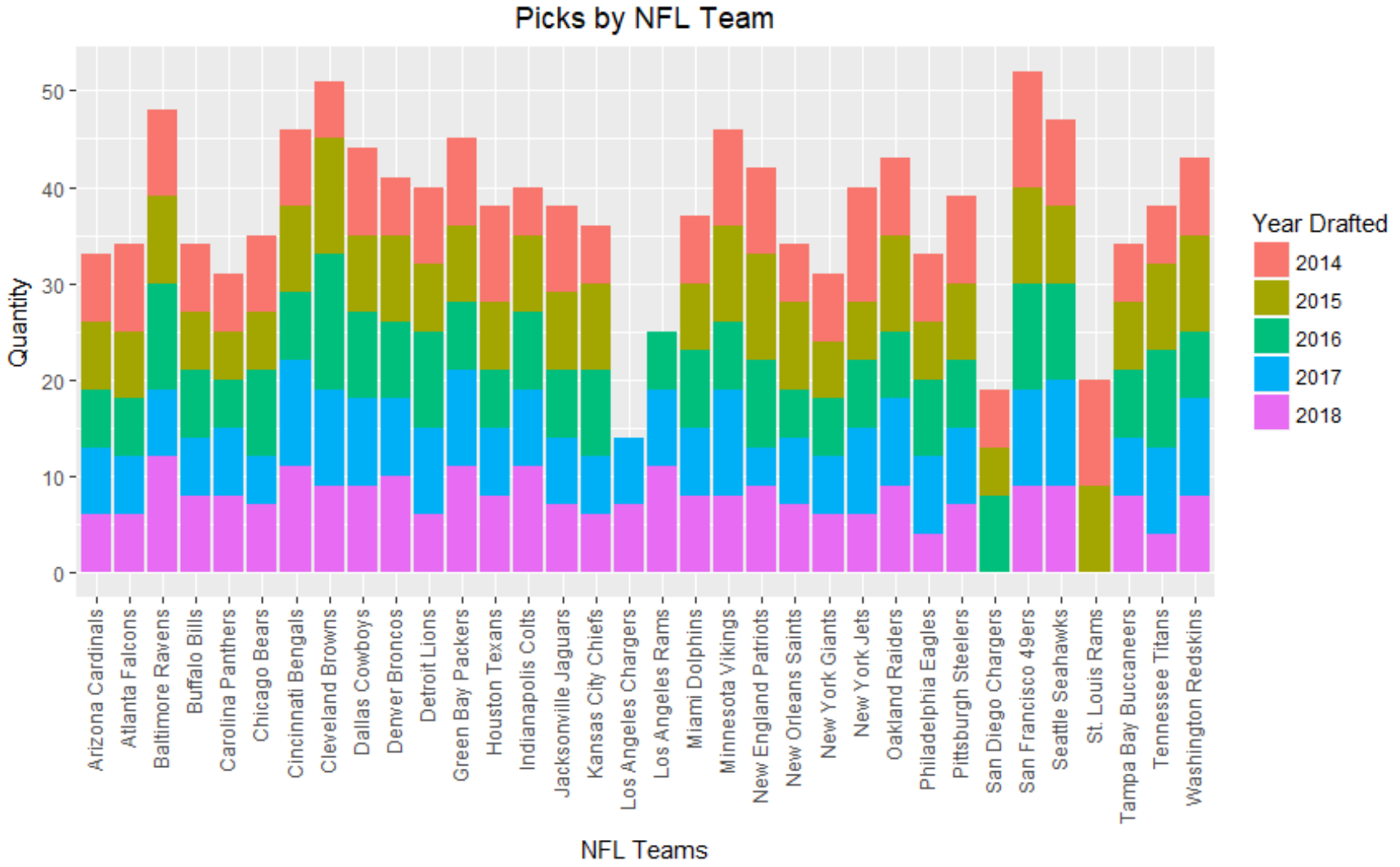
School with most draftees is…Alabama (surprise)

Position player most often picked was WR

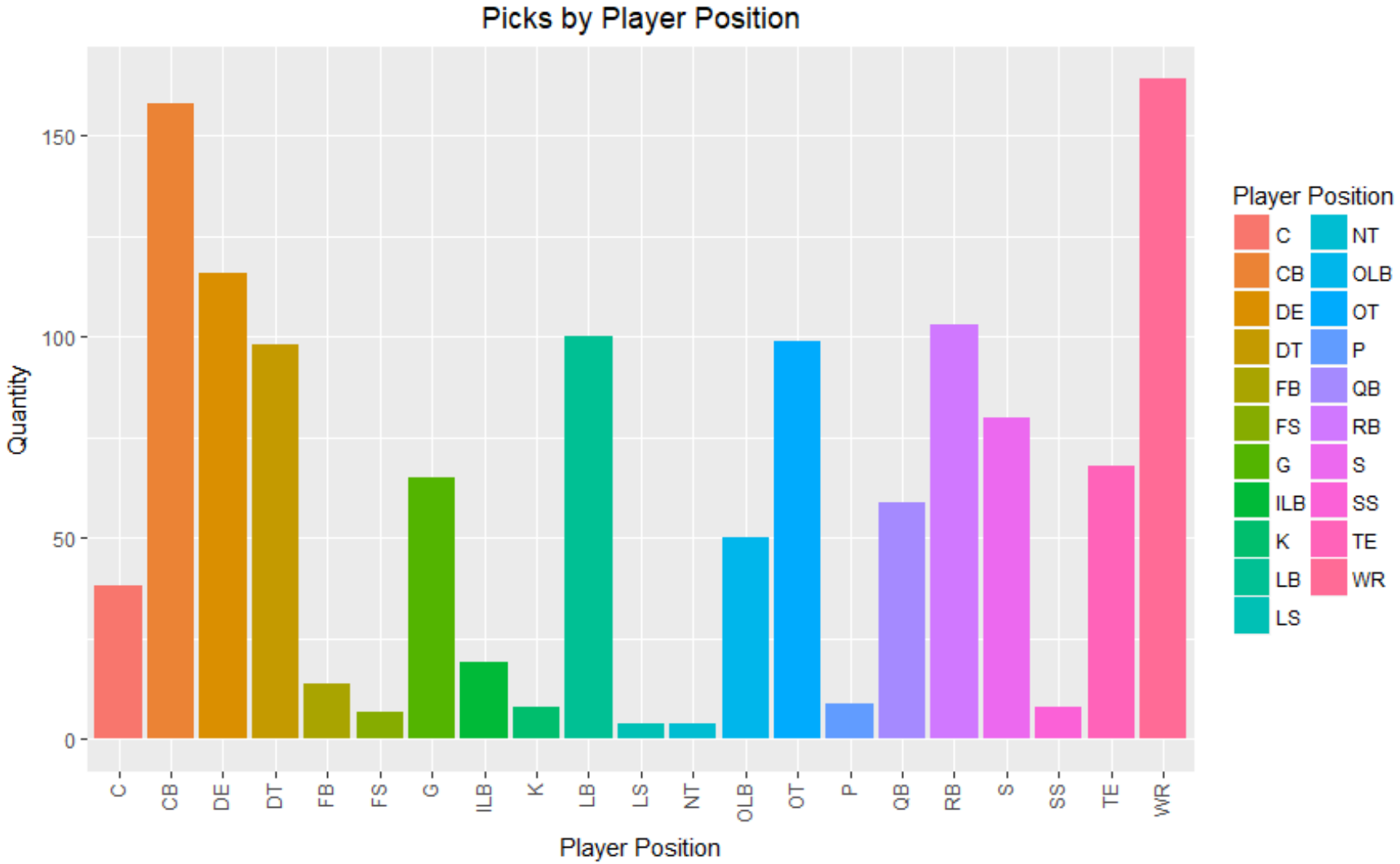
PICKS FROM EACH NCAAF CONFERENCE



PICKS PER NFL TEAM



PICKS BY PLAYER POSITION



ARULES PACKAGE: WHAT DOES IT DO?

CRAN Definition: Provides the infrastructure for representing, manipulating and analyzing transaction data and patterns (frequent item-sets and association rules).

More information about this package can be found on the CRAN website.

For this project: Find associations within draft data. Apriori algorithm finds rules (i.e.: if A happens, then B happens this often)

TERMS TO KNOW

- Support: Number of times a particular item appears as a ratio to the entire number of items in dataset
- Confidence: Ratio of a set of item(s) to only a portion of these items (i.e.: Confidence({A,B}→{C}) = support{A,B,C}/support{A,B})
  - How confident are we that C will be present given that A & B are present.
- Lift: Likelihood that a set of items show up together, 'A, B, & C,' relative to the number of times 'C' will show up across all transactions (or in this case, draft selections).
- Count: Number of times rule items show up together

FINDINGS

Philadelphia Eagles and Carolina Panthers were likely to draft a CB

Green Bay Packers were likely to draft a WR

Buffalo Bills liked to draft from the ACC

Washington Redskins liked to draft from the SEC

All findings and code can be found at github.com/snyousef.