Sports Statistics

Betting Data

Where to get it:
- Odds Portal: http://www.oddsportal.com/
- Betfair: http://data.betfair.com/ - but you have to bet, which at the moment requires not being in the United States
- Fracsoft: http://www.fracsoft.com/ - historical Betfair data, with more details, but extra cost
- Scraping gambling websites:
  - Research the bookmaker. Know what their specialties are.
  - Really depends on the sport.

What to look out for:
- Different types of bookmakers.
  - Peer-vs-peer gambling
  - Traditional bookmakers
- House effect (or the bookmaker's cut or vig)
- When the bets were taken.

Best Bet:
- Talk to people inside of the business.
- Talk to statisticians. (Lots of academics doing work with gambling markets.)
- Talk to trade groups. Both inside and watching the industry.

The Tennis Racket
- https://www.buzzfeed.com/heidiblake/the-tennis-racket
- https://www.buzzfeed.com/johntemplon/how-we-used-data-to-investigate-match-fixing-in-tennis
- https://github.com/BuzzFeedNews/2016-01-tennis-betting-analysis

Rolling Your Own Data

Sometimes you're not going to have the right data for what you want to do, but often whatever body runs the sport releases a lot of valuable public data. You just have to know where to look and how to scrape it.

NCAA: http://stats.ncaa.org/

You can scrape the boxscores and play-by-play for every basketball game. And do things like:

- Calculate how receives each assist from a player:
  http://www.nycbuckets.com/2014/03/breaking-jason-brickmans-1000-assists/
- Figure out how efficient a team is when players are on and off the court:
  http://www.nycbuckets.com/2017/02/nyc-buckets-nec-end-season-awards-conference-teams/

Professional Leagues:

- **Major League Baseball**: http://gd2.mlb.com/components/game/mlb/
- **National Basketball Association:** http://stats.nba.com/
  - Note: This is actually an API. For instance, if you wanted every shot Russell Westbrook has taken this season:
    http://stats.nba.com/stats/shotchartdetail?Period=0&VsConference=&LeagueID=00&LastNGames=0&TeamID=0&PlayerPosition=&Location=&Outcome=&ContextMeasure=FGA&DateFrom=&StartPeriod=&DateTo=&OpponentTeamID=0&ContextFilter=&RangeType=&Season=2016-17&AheadBehind=&PlayerID=201566&EndRange=&VsDivision=&PointDiff=&RookieYear=&GameSegment=&Month=0&ClutchTime=&StartRange=&EndPeriod=&SeasonType=Regular+Season&SeasonSegment=&GameID=
- **Major League Soccer:**
  http://matchcenter.mlssoccer.com/matchcenter/2017-02-23-fc-dallas-vs-arabe-unido/boxscore
  - The boxscores in the GameCenter contain the XY coordinate of every shot taken during the game, which can provide some extra context.
  - It's worth noting that other providers, such as Opta, have even deeper data, with easier access tools, but it costs $$$.

International Organizations:

My colleague Peter Aldhous downloaded data from IAAF and FINA to produce this story: https://www.buzzfeed.com/peteraldhous/faster-higher-stronger-not. You can see the data analysis on Github: https://github.com/BuzzFeedNews/2016-07-athletic-performances