1. **Start with a short introduction**
   a. **Briefly explain what you are going over. Mainly demographics, marketing, and statistical analysis.**
2. **Show that you are only covering a small portion of your work, quickly go through your old slideshow with the extra stuff and mention that it was more than an hour long.**

Demographics area graph
- Explain the key, show how it was expanding then tightened again
- Massive Latino spike
- Latino and African-Americans were sort at equal pace when Latin spiked and continued to grow while the African-American population went way down.
- Asian also started to grow compared to how hard it is to see on the graph

How does this happen:
- Sport changed into a less athletic sport and drew less interest
- Very expensive and hard to get into
- Hard to practice by yourself
- NCAA D1 gives 85 football, 13 Basketball, and 11 baseball gets all to cover around 35 players. Basically not easy to get a scholarship to play baseball.
  - Only 4% of NCAA baseball players were black.

- Also affects the coaching staff, front office, and press
- In 2020 there were only 6 black reporters that mainly covered baseball
- There has been an improvement in all coaching areas

except the manager position
- Used to be 33% in 2009 now down to 20%

- Institute for Diversity and Ethics in Sport (TIDES) at the University of Central Florida makes an MLB racial and gender report card.
  - Racial hiring: B+
  - Gender hiring: C
  - Overall: B
  - Previous grade was B-
  - The league's office is at 37.5% of people of color and 40.1% women, which is similar to the NBA.
  - The card covered the CEO/President for the first time
    - Racial and gender hiring: F
    - 39 of 40 known majority owners are white men
    - GM/Pr.BO got a C-
  - Same issue for women
    - Bright side: Kim Ng became the Marlins hired her as the first female GM ever.
    - 28/30 teams have at least one woman in the role of VP or above.

<mark>Slide 3</mark>
- The same graph as the last just instead this one weights the demographic population by their cumulative WAR to get a better idea of the impact each demographic has.
- WAR: Wins Above Replacement, WAR is one of the most popular stats used in sabermetrics. Its goal is to try and put a number value on a player's performance as a whole. It's not supposed to be incredibly accurate or used on its own like other stats, it's best used in unison with other stats to create a bigger picture. In this situation, it works really well because of how good it is at quantifying player value.
- What mainly stands out is the noticeable increase in size for African Americans compared to the other graph. SImply

what that means is that African American players have been a lot more productive and impactful than the demographic says. Even makes the smaller section from the more recent years look bigger.

- The same goes for Latino players, with a clear increase in size.

<mark>Slide 4</mark>

- Same data as the first demographic graph
- This just makes it easier to see the trend or path of each demographic over time.
- The white population has slowly gone down, which makes sense with the influx of more international players.
- African American was steadily increasing during the 70s and then started to slowly go back down to what looks like a close low to the beginning of the graph.
- Latino spike through the African American population and plateaued a little bit but doesn't seem to be dropping.
- Asian population also starting to go up.

<mark>Slide 5</mark>

- Marketing issue
  - According to SI in 2018 only 22% of the American public knows who Mike Trout is.
  - The best example I can give is that my mom has no idea who he is but know Lebron, Brady, Tiger, Messi, Ronaldo, etc.
  - Trout doesn't do many commercials or sponsorships outside of what the team requires him to do or the league. When you are getting paid like he is you don't really need to do ads. The league like others should be able to market its players.

- He also isn't very flashy or has a very vibrant personality that draws a lot of attention. Unlike some of the current faces of the sport, Vladdy Jr., Tatis Jr., Soto, Acuna Jr.
- Another issue has to do with MLB as a whole, Trout plays for the LA Angels, a team in a large market that somewhat refuses to field a playoff quality team. During one of the only times where viewership is high the league's best player isn't there. Even though playoff viewership is dwindling it's still the best stage to market your game and players. (Trout only has played in 3 playoff games, all in 2014)

- Blackouts
  - You can't watch every local team on tv unless you have the local broadcaster's channel, it's a nationally televised game or you are physically not in the region of that team. Hard for fans to watch their team without paying for an extra streaming service.
- Less interesting draft compared to other professional sports.
  - College baseball doesn't have nearly as much attention as basketball or football. Top college prospects don't get the same national attention that basketball or football prospects do.
  - Because high school players can be drafted it makes it harder to know who is who when watching the draft. A lot harder to take interest in the draft when you don't know anything about the players.
- WS viewer numbers
  - Viewership has gone down which doesn't follow the rest of society.
  - With the growth of the internet and streaming services,

it became easier and more common to watch sports, helping with the growth of most major sports.
- There is a lone peak in 2016 when the Cubs won the World Series, probably the majority were fans from Chicago.
- Otherwise, it's gone down, it's not like it hasn't had major contenders, the Dodgers, Astros, Red Sox, and Mets have been in the WS since 2015.

Three True Outcomes
- TTO represents the current era of baseball. The majority of major league baseball has turned into this, every plate appearance is most likely to end up a walk, strikeout, or home run. Home runs are fun, there's nothing wrong with them but strikeouts are fun and walks are good but it gets old quickly. It was better when you got to see guys showing their athleticism to make great plays on the field instead of waiting for a walk or K.
- This has led a large majority of players into changing their approach in their at-bats. There are more guys who have terrible batting AVGs but have really high OBP and SLG. Stats said that it was good to hit home runs and walk but it ended up making baseball boring.

- This graph shows 2 things for each percentage, the data, and a prediction.
  - Strikeout% is the most obvious, it just shoots up and we already know it's not slowing down anytime soon. 22% is just insane, 1 in 5 PA will be a K, not very enticing to watch. Probably will go up even more and this doesn't even represent the large group of players with 25% K rates.
  - BB% surprisingly has been more steady compared to

the other 2 and hasn't really changed much. It's actually gone down, which makes sense given the influx in strikeouts. Prediction also looks like it will continue to stay steady.
  ○ HR% is a little harder to look at because of how small of a number it is but it's definitely gone up. Primarily given how high it has been 5+ years even compared to the steroid era. That will likely keep going up until there is some change in rules potentially.

Average Fastball Velocity
- This has only been tracked since 2007 but it still shows a lot of change and it's easy to imagine what the number would have been before 07.
- This simply shows the average fastball velocity from 07-21.
- Continuously goes up and doesn't appear to be stopping.
- There is nothing wrong with pitchers throwing harder, but it is part of a wide range of stats that show that pitchers are getting a lot better than hitters with their velocity and pitch design.
  ○ There is so much technology involved with pitching that it makes a lot of sense.
  ○ Higher velo leads to more strikeouts which leads to fewer balls in play and less fun.

Quality of contact %s
- Quality of contact consists of three stats. Soft, medium, and hard contact. It's a known fact in baseball that you have a much better chance of getting a hit it the harder you hit the ball.
- This fuels the three true outcomes, guys just swing out of their shoes all the time and strike out or hit a homerun and some of the time just hit hard lineouts.
- Medium has gone down which makes sense, we are not in the small ball era where there was lots of light contact and

more balls in play.
- Soft has stayed pretty consistent, especially this past decade.
- Hard contact unexpectedly isn't rare anymore and is still going up. Hard % going up or down doesn't have a very clear correlation with the health of baseball. Going up could increase TTO and go down could decrease power. Both are good and bad in their own sense, the game will continue to move in that direction, striving for hard quality contact.

<mark>Slide 9</mark>
Cumulative home runs over the years
- This is one of the better ways to show the power increase even though it's a simple stat.
- Very clearly there is the home run boost in the steroid era, a gentle decline, and then a giant spike up to now with the juiced and not juiced baseball.
- Having the juiced balls in 2019 and maybe before does affect the home run totals but it's hard to give it all the credit. Like I said before guys are swinging out of their shoes more often than not just trying to hit the ball hard and that plus the mentality of trying to hit home runs has played a bigger role in the increase.
- The current totals are above the peak of the steroid era by about 6000 home runs, most of that can just come from the baseball, it's a joint effort.

<mark>Slide 10</mark>
Shift percentages

- One of my favorites because of how drastic it is.
- A shift is when defenders position themselves on the field to where the hitter will most likely hit the ball according to their recent and cumulative spray charts and other stats.
- It's most common for left-handed hitters as the graph shows,

this most likely has to do with lefties being more pull (hit the ball towards your side of the field L-RF, R-LF)happy cumulative
● There is only clear data about this for a short period of time but it's growing very fast. Already in 7 years, there is a large influx in shifts in general but again more lefties.

● The shift is one of the main culprits in making baseball more boring.
● It dampens balls in play and doesn't allow as many hits as there should be.
● Takes away the opportunity for great fielders to shine and show off their athleticism or do their job.
● The goal is to win and to do that you need to score runs and stop runs from being scored on you but this really hampers baseball's ability to be fun to watch.
● Players aren't necessarily losing their athleticism but the game is taking it out.

SLG %

● SLG % is a measurement of all of a batter's extra-base hits. (.400-.450 is pretty good, .450-.500 is very good. Above .500 is great and below .400 is not great)
● SLG being similar to ISO shows a similar picture but SLG has some more inconsistencies.
● Again there is a clear spike in the steroid era than a decline and back up.
● The interesting part is that ISO and HRs are higher now than they have ever been but SLG isn't. That could have to do with the fact that most SLG comes from HRs and not 2B and 3Bs.
● At the moment SLG is down due to weird baseballs and supposedly the short spring training.

- I'm sure it will continue to go back up but I think it won't pass the previous peak as strongly as HRs or ISO did.

Cumulative stolen bases

- As the game has evolved statistics showed that stealing bases wasn't really viable anymore and wasn't worth the risk.
- It's obvious that having a runner on second is very good for a lot of reasons, negates a double play, easier to score off a single to the outfield from 2nd, and distracts the pitcher.
- But the higher potential to be tagged out and the chance of injuring your hand mitigate the benefits of stealing bases.
- There's a reason Mike Trout once led the league in stolen bases but now barely steals any, the injuries you could sustain in your fingers or hand from stealing bases can shut down a player for a while and are the main cause of the decrease in stolen bases.
- The graph peeks in the late 90s and the late 70s but has consistently gone down since then.
- As of now, the total is down a little more than 3000 SBs compared to the highest peak in the late 90s.

ISO graph
- ISO basically gives an image of a player's raw power and how many extra bases they get per at-bat.
- ISO means isolated power, it's similar to slugging in it just looks at a player's power and no other factors.
- ISO can help differentiate between stats that are closer together. Having a similar AVG won't correlate to OBP or SLG.
- It's hard to see the growth but similar to HR% the changes

are in small increments that matter more than they look.
- Even with the small increases, the past 6 years AVG is higher than 1 hundredth above the last peak in the late 90s early 00s.
- Most likely will continue to go up along with other power stats.

- First thing that was really important in going through the project was my timing.
  - I tried my best to plan out what i wanted to do each day but you can only go so far ahead.
  - Sometimes I didn't do as much as i wanted one day or couldn't finish something i wanted to in one day.
  - The main part is that the time goes by very quickly, the first part of that is the realization that you dont get all of May for the project but about ¾ and thats deceiving too. Basically you have 2 1/2 weeks to get your work done then the rest is for making a presentation or whatnot.
- Really settling on what I wanted to do was tough, at first before i even started i originally planned on trying to make a prediction model and further expand that into other interesting statistical analysis topics in baseball.
  - I had been trying to make a model for a few months already and I couldn't really get it right so i decided that I was better off going more research style then trying to push myself to do something I might not be able to do.
  - I settled on doing research on a few different topics, financials, tv contracts & market size, baseball demographics, marketing, MiLB issues, and some statistical questions i wanted to look at.

- I started by spending about a week or so going through maybe 30 articles/sources taking 20+ pages of notes if not more. You do not need to do this neither did I, i thought i was going to need way more information to make a presentation long enough but i was very wrong.
- I then spent some time gathering information from my notes for my slideshow and making some graphs and tables from the data that I had already collected.
- From there I moved onto my favorite but most challenging section.
  - I needed to gather the data I had found in some form, most of the data i got for my financials, tv contracts, and market size where small enough to hand copy.
  - Most of the data i was looking at was mainly designed to be looked at in an easy fashion for fans and not really designed to be heavily manipulated.
  - Starting with my demographic data, the hardest one. I had the location of the data but no way to download it and it was too long to hand copy, so i taught myself how to code a webscraper that i used to scrape the data off the page, turned it into a dataframe, then some graphs.
  - For my data on statistical questions it was a little easier but more tedious. Most of them i was able to get the data from FanGraphs or BaseballReference, and sometimes Statcast. For these sites you are able to download the data but you can only do so much with that. So for each stat I made short scripts to get the data from the website for the required timeframe, clean the data, then set it up for export where i would then make the graphs i showed you in excel. Most of the

coding I did was more tedious and time consuming then challenging but I did have my fair share of struggles. Mainly because i couldn't just reuse my code for each stat because i was manipulating them differently for different time periods and some were aggregate some were single season so i had to custom tailor my scripts for about 5+ different setups.

- Making the presentation was byfar my least favorite part, it was really tricky trying to compress all my information.
- Partly becuase i hate presentations and partly because i was sad i didn't have any more data gathering or coding left to do.

Takeaways

- Mostly what i learned is that if I'm legitimately interested in something I can put a lot more effort and time into it than i thought i could, i think this would apply to most people obviously.
- I added to my understanding of baseballs issues but i realized that as a sport compared to our society and other major sports and even other countries its gone backwards. I can't really think of anything where diversity has sort of gone backwards in a weird way, excluding some and welcoming others for not very clear reasons. At the end of the day a good ball player is a good ball player.
- On top of that i learned that data really is everywhere, its captured for almost everything now, both for bad and good intents. I do think that there is a lot of good that can be done with data and this project really helped solidify two things for me. 1. Is that i still know that i would love to work in a major

league baseball front office. 2. I'm even more set on my major of data science then before, after having real experience working with data and seeing all the interesting things you can do with it. I'm incredibly excited to get more involved and see what I will do in the future.