Moneyball: The Role of Socioeconomic Status in the Changing Demographic of NCAA Division I Basketball

Jasmine Sanders
University of California, Berkeley
Sociology, PhD

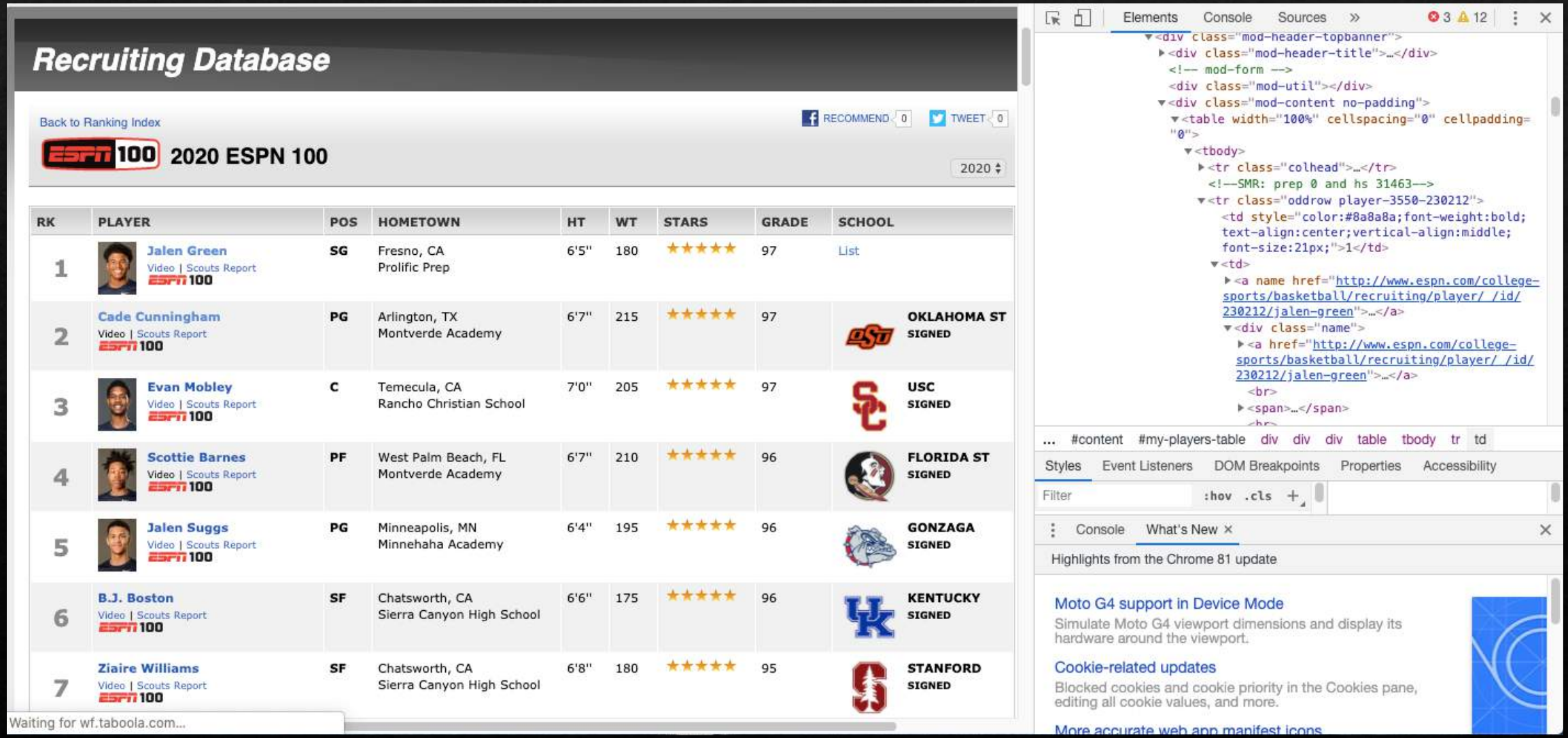
Context & Research Questions

- First-generation college students in Division I sports fell from 16% in 2010 to 14% in 2015[1]
  - Percentage of Division I men's basketball players declined by a third from 28% to 19%

1. What contextual factors contribute to the decline in signing first-generation college athletes on Division I Men's Basketball teams?
   1. To what extent does private/elite school attendance impact this shift?

[1] NCAA GOALS, 2011 & 2015

ESPN 100 Recruiting Database



Recruiting Database

Back to Ranking Index

RECOMMEND 0 TWEET 0

ESPN 100 2020 ESPN 100

2020

RK	PLAYER	POS	HOMETOWN	HT	WT	STARS	GRADE	SCHOOL
1	Jalen Green Video | Scouts Report ESPN 100	SG	Fresno, CA Prolific Prep	6'5"	180	★★★★★	97	List
2	Cade Cunningham Video | Scouts Report ESPN 100	PG	Arlington, TX Montverde Academy	6'7"	215	★★★★★	97	OKLAHOMA ST SIGNED
3	Evan Mobley Video | Scouts Report ESPN 100	C	Temecula, CA Rancho Christian School	7'0"	205	★★★★★	97	USC SIGNED
4	Scottie Barnes Video | Scouts Report ESPN 100	PF	West Palm Beach, FL Montverde Academy	6'7"	210	★★★★★	96	FLORIDA ST SIGNED
5	Jalen Suggs Video | Scouts Report ESPN 100	PG	Minneapolis, MN Minnehaha Academy	6'4"	195	★★★★★	96	GONZAGA SIGNED
6	B.J. Boston Video | Scouts Report ESPN 100	SF	Chatsworth, CA Sierra Canyon High School	6'6"	175	★★★★★	96	KENTUCKY SIGNED
7	Ziaire Williams Video | Scouts Report ESPN 100	SF	Chatsworth, CA Sierra Canyon High School	6'8"	180	★★★★★	95	STANFORD SIGNED

Waiting for wf.taboola.com...

Elements Console Sources » 3 12

```
<div class="mod-header-topbanner">
  <div class="mod-header-title">…</div>
  <!-- mod-form -->
  <div class="mod-util"></div>
  <div class="mod-content no-padding">
    <table width="100%" cellspacing="0" cellpadding="0">
      <tbody>
        <tr class="colhead">…</tr>
        <!--SMR: prep 0 and hs 31463-->
        <tr class="oddrow player-3550-230212">
          <td style="color:#8a8a8a;font-weight:bold;text-align:center;vertical-align:middle;font-size:21px;">1</td>
          <td>
            <a name href="http://www.espn.com/college-sports/basketball/recruiting/player/_/id/230212/jalen-green">…</a>
            <div class="name">
              <a href="http://www.espn.com/college-sports/basketball/recruiting/player/_/id/230212/jalen-green">…</a>
              <br>
              <span>…</span>
```

... #content #my-players-table div div div table tbody tr td

Styles Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls +

Console What's New ×

Highlights from the Chrome 81 update

Moto G4 support in Device Mode

Simulate Moto G4 viewport dimensions and display its hardware around the viewport.

Cookie-related updates

Blocked cookies and cookie priority in the Cookies pane, editing all cookie values, and more.

More accurate web app manifest icons

Script Highlights

```
for year in range(2010,2021):
    print(year)
    url='http://www.espn.com/college-sports/basketball/recruiting/playerrankings/_/class/'+str(year)+'/order/true'
    print(url)

    try:
        page = urlopen('http://www.espn.com/college-sports/basketball/recruiting/playerrankings/_/class/'+str(year)+'/order/true')
    except HTTPError as e:
        print(e) # The HTTP error: "404 Page Not Found (you messd up)" or 500 Internal Server Error (I messed up)"
    except URLError as e:
        print("The server is broken") # No server could be reached
    else:
        print("The site is working")
#
#_____________________________________________________________________________

#Insert beautiful soup code
    from bs4 import BeautifulSoup

    soup = BeautifulSoup(page, "html.parser")

#_____________________________________________________________________________
### Title of the page can be displayed by .title
    print(soup.title)
    type(soup.title)

    # ### You can get text by adding .text
    print(soup.title.text)
    type(soup.title.text)



#_____________________________________________________________________________
### Extracting a table
    #Pattern find_all(tagName, tagAttributes)
    roster_even_table = soup.findAll('tr',
                                     {'class':'evenrow'})
    print(roster_even_table)

    roster_odd_table = soup.findAll('tr',
                                     {'class':'oddrow'})
    print(roster_odd_table)

#
```

```
##NOW CODE FOR DATA SET
##Code to take bio and write it to a file
##or just keep adding to a dataframe
all_even_urls=[]

for i in range(0,50):
        #print(i)
        url = roster_even_table[i].find('a').get('href')
        print(url)
        all_even_urls.append(url)

print(all_even_urls)

all_odd_urls=[]
for i in range(0,50):
        #print(i)
        url=roster_odd_table[i].find('a').get('href')
        print(url)
        all_even_urls.append(url)
```

```
bio_list = []
hometown = []
school = []
position = []
status = []

for url in all_even_urls:
    req = urlopen(url)
    page = BeautifulSoup(req, 'html.parser')
    bio = page.find('div', # find element
                    {'class':'bio'})
    for a in bio.findAll('a'):
        a.replaceWithChildren()
    for section in bio.findAll('ul'):
        cells = section.findAll(['li']) # to iterater through each row
    if len(cells) == 4: # no heading
        hometown.append(cells[0].find_all(text=re.compile('[A-Z]+'))) # there's
        school.append(cells[1].find_all(text=re.compile('[A-Z]+')))
        position.append(cells[2].find(text=re.compile('[A-Z]+')))
        status.append(cells[3].find(text=re.compile('[A-Z]+')))
    else:
        print("something is wrong") # for debugging
    bio_list.append(bio)

print(hometown)
print(school)

import pandas as pd # convention

div_pd = pd.DataFrame() # create a data frame
div_pd['hometown'] = hometown
div_pd['high school'] = school

print(div_pd)

div_pd.to_csv("final_project_2020.csv")
```

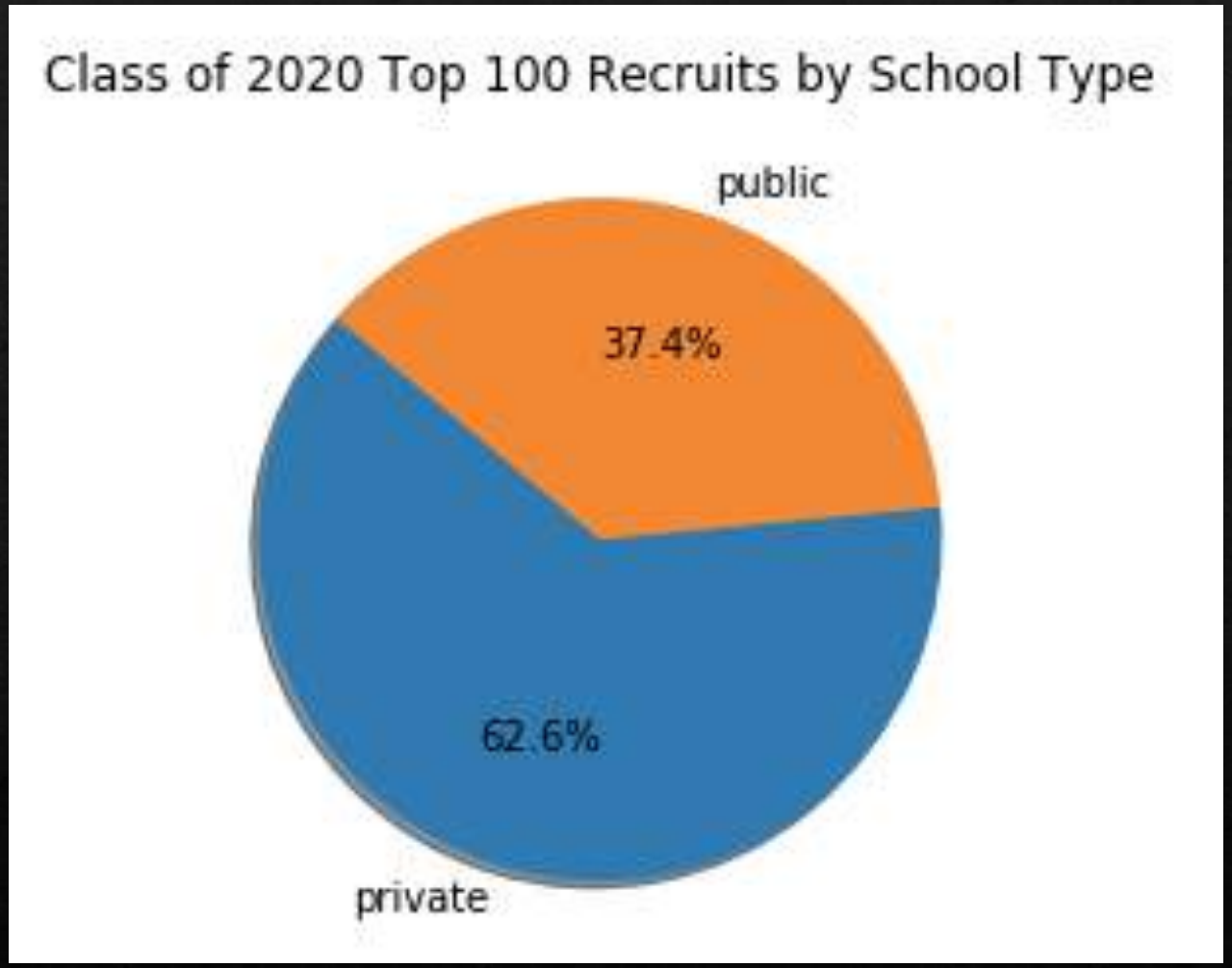
Excel Data

A	B	C	D	E
year		hometown	school	type
2020	0	Arlington, Texas	Montverde Academy	private
2020	1	West Palm Beach, Fla.	Montverde Academy	private
2020	2	Chatsworth, Calif.	Sierra Canyon High School	private
2020	6	Richmond, Va.	Word of God Christian Academy	private
2020	7	Charlotte, N.C.	IMG Academy	private
2020	8	Leesburg, Va.	Paul VI High School	private
2020	9	Las Vegas, Nev.	Trinity International School	private
2020	10	Chesapeake, Va.	Oak Hill Academy	private
2020	11	Santa Ana, Calif.	Mater Dei High School	private
2020	13	Jackson, Miss.	Huntington Prep	private
2020	15	Albany, N.Y.	The Albany Academy	private
2020	16	Hyattsville, Md.	DeMatha Catholic High School	private
2020	17	Memphis, Tenn.	IMG Academy	private
2020	18	Hyattsville, Md.	DeMatha Catholic High School	private
2020	19	White Plains, N.Y.	Archbishop Stepinac High School	private
2020	20	Little Rock, Ark.	Montverde Academy	private
2020	21	Seattle, Wash.	Dream City Christian	private
2020	24	Montclair, N.J.	Roselle Catholic High School	private
2020	26	Bradenton, Fla.	IMG Academy	private
2020	27	Murrieta, Calif.	Rancho Christian School	private
2020	29	Tempe, Ariz.	Hillcrest Prep	private
2020	34	Decatur, Ga.	Hargrave Military Academy	private
2020	35	Las Vegas, Nev.	Huntington Prep	private
2020	36	Toledo, Ohio	Montverde Academy	private
2020	37	Alpharetta, Ga.	St. Francis High School	private
2020	38	Newark, N.J.	Roselle Catholic High School	private
2020	43	Brooklyn, N.Y.	Brewster Academy	private

```
import xlrd
import matplotlib.pyplot as plt
import numpy as np


df = pd.read_csv("/Users/jasminesanders/Final Project/final_project_2020_graph.csv")
school_data = df.groupby("type")
types = ['private', 'public']
school_data.size()
colors = ["#1f77b4", "#ff7f0e"]
plt.pie(school_data.size(), labels=types, colors=colors,
autopct='%1.1f%%', shadow=True, startangle=140)
plt.title("Class of 2020 Top 100 Recruits by School Type")
plt.show()
plt.savefig('Final_project_pie chart.pdf')

#
```


Class of 2020 Top 100 Recruits by School Type
public
37.4%
62.6%
private

Future Directions

Continue Year Loop

Combine Roster Data + School Type

Show trends across decade (2010-2020)