#### **Project**

CPSC 260

This project will require you to use the skills that you have learned throughout the semester. You will be performing some simple statistical analysis (data summaries and visualization) on data that you scrape from the web. The data will need to be cleaned or pre-processed before you can use it.

You will turn in a copy of your annotated R script along with this document. Answer each of the questions below as you work.

**Data Collection Requirements:** You must use at least two webpages, and you must scrape data for at least 5 variables total (between the different webpages). At least two should be categorical variables, and at least 3 should be quantitative variables. This means you must collect both numeric and character data. Keep this in mind as you answer the first few questions!

#### Choosing a Topic and Collecting Data:

1) Come up with a topic to research and explain why it is relevant to you.

**Topic: NBA stats** 

Relevance: I enjoy playing and watching basketball and I think it would be interesting to dig more into the stats nobody looks at and also some of the more common stats that people also debate over

- 2) Brainstorm at least seven possible research questions related to your topic that you would like to explore by scraping data from the web.
  - 1. The most common state/city NBA players are born in
  - 2. How many NBA players have the same name or at least first name
  - 3. Average age of current NBA players
  - 4. Points per game rankings for current players
  - 5. Average salary
  - 6. Assists per game rankings for current players
  - 7. How assists vs Points vs Salary compare
  - 8. What college they attended
- 3) Find multiple web pages containing data that will help you explore as many of your questions from #2 as possible. Once you find appropriate webpages, read them into R and perform your web scraping. Provide a brief description here of each web page and what information you scraped from each. [See data collection requirements above!]

Webpage #1 Description and Scraped Data:

# NBA & ABA Career Leaders and Records for Points Per Game | Basketball-Reference.com

- NBA Points per game leaders for current players

Webpage #2 Description and Scraped Data:

# NBA & ABA Career Leaders and Records for Assists Per Game | Basketball-Reference.com

- NBA Assists per game leaders for current players

Webpage #3 Description and Scraped Data:

#### NBA Salary Rankings | Spotrac

- NBA salary rankings for current players

Webpage #4 Description and Scraped Data:

#### NBA Players & Team Rosters | NBA.com

- NBA player birthplace and college

#### Preparing your data for use:

- 4) Clean up your data so that it is ready for analysis and visualization. As you work describe what you have done at each step. Along with any other work that needs to be done, make sure to cover the following:
  - a) Convert variables/data into the correct format.
  - b) Check for outliers and errors. If you find "problems", decide how to handle them.
  - c) Check for missing values. If you find missing values, decide how to handle them.
  - d) Perform any necessary data transformations.

This is similar to the data pre-processing homework. Also see the R script for the web scraping examples.

Detailed list of steps taken to perform data-preprocessing/ data cleaning: Bring in all 4 websites using read\_html and make the respective stat into a table. Added a rank column to 'SalaryTable2' Because there was no column name. Changed the 5<sup>th</sup> column name to 'Salary'. Only selecting rank, player, and salary from the salary website and filling in the NA rank values

using [row number, column number]. Fixed All of the player names in salary too because they repeated for some reason.

	Player	Position <sup>‡</sup>	Age <sup>‡</sup>	cap hit
1	Curry Stephen Curry GSW	PG	33	\$45,780,966
2	Harden James Harden PHI	SG	32	\$44,310,840
NA	Wall John Wall HOU	PG	31	\$44,310,840
4	Westbrook Russell Westbrook LAL	PG	33	\$44,211,146
5	Durant Kevin Durant BKN	SF	33	\$42,018,900
6	James LeBron James LAL	SF	37	\$41,180,544
7	George Paul George LAC	SG	31	\$39,344,900
NA	Leonard Kawhi Leonard LAC	SF	30	\$39,344,900
NA	Lillard Damian Lillard POR	PG	31	\$39,344,900
NA	Antetokounmpo Giannis Antetokounmpo MIL	PF	27	\$39,344,900
11	Thompson Klay Thompson GSW	SG	31	\$37,980,720
12	Butler Jimmy Butler MIA	SF	32	\$36,016,200
13	Harris Tobias Harris PHI	SF	29	\$35,995,950
14	Middleton Khris Middleton MIL	SF	30	\$35,500,000
15	Davis Anthony Davis LAL	PF	28	\$35,361,360
16	Gobert Rudy Gobert UTH	С	29	\$35,344,828
17	Irvina Kvrie Irvina BKN	PG	29	\$35 328 700

^	Rank <sup>‡</sup>	Player	Position <sup>‡</sup>	Age <sup>‡</sup>	Salary <sup>‡</sup>
1	1	Curry Stephen Curry GSW	PG	33	\$45,780,966
2	2	Harden James Harden PHI	SG	32	\$44,310,840
3	NA	Wall John Wall HOU	PG	31	\$44,310,840
4	4	Westbrook Russell Westbrook LAL	PG	33	\$44,211,146
5	5	Durant Kevin Durant BKN	SF	33	\$42,018,900
6	6	James LeBron James LAL	SF	37	\$41,180,544
7	7	George Paul George LAC	SG	31	\$39,344,900
8	NA	Leonard Kawhi Leonard LAC	SF	30	\$39,344,900
9	NA	Lillard Damian Lillard POR	PG	31	\$39,344,900
10	NA	Antetokounmpo Giannis Antetokounmpo MIL	PF	27	\$39,344,900
11	11	Thompson Klay Thompson GSW	SG	31	\$37,980,720
12	12	Butler Jimmy Butler MIA	SF	32	\$36,016,200
13	13	Harris Tobias Harris PHI	SF	29	\$35,995,950
14	14	Middleton Khris Middleton MIL	SF	30	\$35,500,000
15	15	Davis Anthony Davis LAL	PF	28	\$35,361,360
16	16	Gobert Rudy Gobert UTH	С	29	\$35,344,828

^	Rank <sup>‡</sup>	Player	Salary
1	1	Curry Stephen Curry GSW	\$45,780,966
2	2	Harden James Harden PHI	\$44,310,840
3	3	Wall John Wall HOU	\$44,310,840
4	4	Westbrook Russell Westbrook LAL	\$44,211,146
5	5	Durant Kevin Durant BKN	\$42,018,900
6	6	James LeBron James LAL	\$41,180,544
7	7	George Paul George LAC	\$39,344,900
8	8	Leonard Kawhi Leonard LAC	\$39,344,900
9	9	Lillard Damian Lillard POR	\$39,344,900
10	10	Antetokounmpo Giannis Antetokounmpo MIL	\$39,344,900
11	11	Thompson Klay Thompson GSW	\$37,980,720
12	12	Butler Jimmy Butler MIA	\$36,016,200
13	13	Harris Tobias Harris PHI	\$35,995,950
14	14	Middleton Khris Middleton MIL	\$35,500,000
15	15	Davis Anthony Davis LAL	\$35,361,360
16	16	Gobert Rudy Gobert UTH	\$35,344,828
17	17	Irvina Kvrie Irvina BKN	\$35,328,700

Use inner\_join to join PPG, APG, and salary by player name.

^	Rank <sup>‡</sup>	Player	PPG <sup>‡</sup>
1	1	Kevin Durant	27.18
2	2	LeBron James	27.13
3	3	James Harden	24.92
4	4	Damian Lillard	24.63
5	5	Stephen Curry	24.29
6	6	Anthony Davis	23.82
7	7	Devin Booker	23.52
8	8	Karl-Anthony Towns	23.17
9	9	Kyrie Irving	23.06
10	10	Russell Westbrook	22.82
11	11	Carmelo Anthony	22.45
12	12	Bradley Beal	22.06
13	13	Giannis Antetokounmpo	21.83
14	14	DeMar DeRozan	20.76
15	15	Paul George	20.39
16	16	Blake Griffin	19.81
17	17	7ach LaVine	19.80

_	Rank <sup>‡</sup>	Player	APG <sup>‡</sup>
1	1	Chris Paul	9.50
2	2	John Wall	9.07
3	3	Russell Westbrook	8.43
4	4	Rajon Rondo	7.92
5	5	Ricky Rubio	7.59
6	6	LeBron James	7.35
7	7	James Harden	6.79
8	8	Damian Lillard	6.65
9	9	Stephen Curry	6.52
10	10	Jrue Holiday	6.40
11	11	Kyle Lowry	6.32
12	12	Nikola Jokic	6.23
13	13	Elfrid Payton	5.74
14	14	Kyrie Irving	5.73
15	15	Mike Conley	5.64
16	16	D'Angelo Russell	5.59
17	17	Jeff Teague	5.55

^	Rank.x <sup>‡</sup>	Player	PPG <sup>‡</sup>	Rank.y <sup>‡</sup>	APG <sup>‡</sup>
1	1	Kevin Durant	27.18	32	4.29
2	2	LeBron James	27.13	6	7.35
3	3	James Harden	24.92	7	6.79
4	4	Damian Lillard	24.63	8	6.65
5	5	Stephen Curry	24.29	9	6.52
6	6	Anthony Davis	23.82	83	2.34
7	7	Devin Booker	23.52	28	4.67
8	8	Karl-Anthony Towns	23.17	63	3.08
9	9	Kyrie Irving	23.06	14	5.73
10	10	Russell Westbrook	22.82	3	8.43
11	11	Carmelo Anthony	22.45	75	2.72
12	12	Bradley Beal	22.06	35	4.19
13	13	Giannis Antetokounmpo	21.83	29	4.60
14	14	DeMar DeRozan	20.76	41	3.90
15	15	Paul George	20.39	50	3.58
16	16	Blake Griffin	19.81	36	4.14
17	17	7ach LaVine	19.80	43	3.89

•	Rank.x <sup>‡</sup>	Player	PPG <sup>‡</sup>	Rank.y <sup>‡</sup>	APG <sup>‡</sup>	Rank <sup>‡</sup>	Salary <sup>‡</sup>
1	1	Kevin Durant	27.18	32	4.29	5	\$42,018,900
2	2	LeBron James	27.13	6	7.35	6	\$41,180,544
3	3	James Harden	24.92	7	6.79	2	\$44,310,840
4	4	Damian Lillard	24.63	8	6.65	9	\$39,344,900
5	5	Stephen Curry	24.29	9	6.52	1	\$45,780,966
6	6	Anthony Davis	23.82	83	2.34	15	\$35,361,360
7	7	Devin Booker	23.52	28	4.67	23	\$31,650,600
8	8	Karl-Anthony Towns	23.17	63	3.08	24	\$31,650,600
9	9	Kyrie Irving	23.06	14	5.73	17	\$35,328,700
10	10	Russell Westbrook	22.82	3	8.43	4	\$44,211,146
11	12	Bradley Beal	22.06	35	4.19	18	\$33,724,200
12	13	Giannis Antetokounmpo	21.83	29	4.60	10	\$39,344,900
13	14	DeMar DeRozan	20.76	41	3.90	41	\$26,000,000
14	15	Paul George	20.39	50	3.58	7	\$39,344,900
15	17	Zach LaVine	19.80	43	3.89	58	\$19,500,000
16	20	Klay Thompson	19.55	85	2.32	11	\$37,980,720
17	าา	Androw Wingsins	10.26	0.4	2 22	27	\$21 E70 200

Brought in a new website that included the player birthplace and college but the names did not have a space for some reason so I removed the space from the player names in the data I merged before this and then inner\_join that data by 'PlayerName'.

^	Player	Team 🗘	Number	Position <sup>‡</sup>	Height <sup>‡</sup>	Weight <sup>‡</sup>	Last Attended	Country
1	PreciousAchiuwa	TOR	5	F	6-8	225 lbs	Memphis	Nigeria
2	StevenAdams	MEM	4	С	6-11	265 lbs	Pittsburgh	New Zea
3	BamAdebayo	MIA	13	C-F	6-9	255 lbs	Kentucky	USA
4	SantiAldama	MEM	7	F-C	6-11	215 lbs	Loyola-Maryland	Spain
5	LaMarcusAldridge	BKN	21	C-F	6-11	250 lbs	Texas-Austin	USA
6	NickeilAlexander-Walker	UTA	6	G	6-5	205 lbs	Virginia Tech	Canada
7	GraysonAllen	MIL	7	G	6-4	198 lbs	Duke	USA
8	JarrettAllen	CLE	31	С	6-10	243 lbs	Texas-Austin	USA
9	JoseAlvarado	NOP	15	G	6-0	179 lbs	Georgia Tech	USA
10	JustinAnderson		NA	F-G	6-5	231 lbs	Virginia	USA
11	KyleAnderson	MEM	1	F-G	6-9	230 lbs	UCLA	USA
12	Giannis Antetoko unmpo	MIL	34	F	6-11	242 lbs	Filathlitikos	Greece
13	ThanasisAntetokounmpo	MIL	43	F	6-6	219 lbs	Panathinaikos	Greece
14	CarmeloAnthony	LAL	7	F	6-7	238 lbs	Syracuse	USA
15	ColeAnthony	ORL	50	G	6-3	185 lbs	North Carolina	USA
16	OGAnunoby	TOR	3	F	6-7	232 lbs	Indiana	United K

^	Rank.x	Player	PPG ÷	Rank.y	APG <sup>‡</sup>	Rank <sup>‡</sup>	Salary <sup>‡</sup>	PlayerName
1	1	Kevin Durant	27.18	32	4.29	5	\$42,018,900	KevinDurant
2	2	LeBron James	27.13	6	7.35	6	\$41,180,544	LeBronJames
3	3	James Harden	24.92	7	6.79	2	\$44,310,840	JamesHarden
4	4	Damian Lillard	24.63	8	6.65	9	\$39,344,900	DamianLillard
5	5	Stephen Curry	24.29	9	6.52	1	\$45,780,966	StephenCurry
6	6	Anthony Davis	23.82	83	2.34	15	\$35,361,360	AnthonyDavis
7	7	Devin Booker	23.52	28	4.67	23	\$31,650,600	DevinBooker
8	8	Karl-Anthony Towns	23.17	63	3.08	24	\$31,650,600	Karl-AnthonyTowns
9	9	Kyrie Irving	23.06	14	5.73	17	\$35,328,700	Kyrielrving
10	10	Russell Westbrook	22.82	3	8.43	4	\$44,211,146	RussellWestbrook
11	12	Bradley Beal	22.06	35	4.19	18	\$33,724,200	BradleyBeal
12	13	Giannis Antetokounmpo	21.83	29	4.60	10	\$39,344,900	Giannis Antetoko unmpo
13	14	DeMar DeRozan	20.76	41	3.90	41	\$26,000,000	DeMarDeRozan
14	15	Paul George	20.39	50	3.58	7	\$39,344,900	PaulGeorge
15	17	Zach LaVine	19.80	43	3.89	58	\$19,500,000	ZachLaVine
16	20	Klay Thompson	19.55	85	2.32	11	\$37,980,720	KlayThompson
(								>

Changed "Last attended" to "College" and then only selected the 'PlayerName', 'College', and 'Country' Columns.

^	PlayerName	College <sup>‡</sup>	Country
1	PreciousAchiuwa	Memphis	Nigeria
2	StevenAdams	Pittsburgh	New Zealand
3	BamAdebayo	Kentucky	USA
4	SantiAldama	Loyola-Maryland	Spain
5	LaMarcusAldridge	Texas-Austin	USA
6	NickeilAlexander-Walker	Virginia Tech	Canada
7	GraysonAllen	Duke	USA
8	JarrettAllen	Texas-Austin	USA
9	JoseAlvarado	Georgia Tech	USA
10	JustinAnderson	Virginia	USA
11	KyleAnderson	UCLA	USA
12	Giannis Antetokounmpo	Filathlitikos	Greece
13	ThanasisAntetokounmpo	Panathinaikos	Greece
14	CarmeloAnthony	Syracuse	USA
15	ColeAnthony	North Carolina	USA
16	OGAnunoby	Indiana	United Kingdom
17	RvanArcidiacono	Villanova	USA

Inner\_join the data with PPG, APG, and Salary with this data. The URL for the website I used wasn't changing when I was selecting the player page, so it only shows players with last names A and B. The website didn't have the whole table on one page so you had to click through pages but the URL never changed, therefore it wouldn't allow me to get the rest of the player names or it all would've looked like the pictures below.

^	Playe	rName	\$	College	\$	Count	у 🗦	Rank.x	\$	Player		<b>‡</b>	PPG	0	Rank	.y <sup>‡</sup>	AP	G <sup>‡</sup>	Rank
1	Giann	isAntetokounm	ро	Filathlitiko	S	Greece	!	1	13	Giannis	s Antetoko	unmpo	2	1.83		29		4.60	
2	WillBa	rton		Memphis		USA		g	92	Will Ba	rton		1	1.63		74		2.74	
3	Bradle	eyBeal		Florida		USA		1	12	Bradle	y Beal		2	2.06		35		4.19	
e	\$	College <sup>‡</sup>	Co	untry <sup>‡</sup>	Rai	nk.x	<sup>‡</sup> Pla	yer		÷	PPG <sup>‡</sup>	Rank.	<b>y</b>	APO	÷	Rank	÷	Salary	, ÷
tokoı	unmpo	Filathlitikos	Gr	eece		1	3 Gia	annis Anteto	kou	unmpo	21.83		29		4.60		10	\$39,34	44,900
		Memphis	US	Α		9	2 Wi	ll Barton			11.63		74		2.74		82	\$15,6	25,000
		Florida	US	Α		1	2 Bra	adley Beal			22.06		35		4.19		18	\$33,7	24,200

Instead of trying to work with 3 names, I used the data before I included the college and country. I removed one of the 'Player' columns since It repeated player names twice. Reordered the columns to put the name first. Then renamed the 'Rank.x', 'Rank', and 'Rank.y' to correspond to their stat and have a meaningful name.

^	Rank.x	PPG ÷	Rank.y	APG <sup>‡</sup>	Rank	Salary	PlayerName
1	1	27.18	32	4.29	5	\$42,018,900	KevinDurant
2	2	27.13	6	7.35	6	\$41,180,544	LeBronJames
3	3	24.92	7	6.79	2	\$44,310,840	JamesHarden
4	4	24.63	8	6.65	9	\$39,344,900	DamianLillard
5	5	24.29	9	6.52	1	\$45,780,966	StephenCurry
6	6	23.82	83	2.34	15	\$35,361,360	AnthonyDavis
7	7	23.52	28	4.67	23	\$31,650,600	DevinBooker
8	8	23.17	63	3.08	24	\$31,650,600	Karl-AnthonyTowns
9	9	23.06	14	5.73	17	\$35,328,700	Kyrielrving
10	10	22.82	3	8.43	4	\$44,211,146	RussellWestbrook
11	12	22.06	35	4.19	18	\$33,724,200	BradleyBeal
12	13	21.83	29	4.60	10	\$39,344,900	GiannisAntetokounmpo
13	14	20.76	41	3.90	41	\$26,000,000	DeMarDeRozan
14	15	20.39	50	3.58	7	\$39,344,900	PaulGeorge
15	17	19.80	43	3.89	58	\$19,500,000	ZachLaVine
16	20	19.55	85	2.32	11	\$37,980,720	KlayThompson
17	22	19.26	84	2.33	27	\$31 579 390	AndrewWiggins

•	PlayerName <sup>‡</sup>	Salary <sup>‡</sup>	Rank <sup>‡</sup>	PPG	Rank	. <b>х</b> ‡	APG	<sup>‡</sup> Rank.y	÷	
1	KevinDurant	\$42,018,900	5	27.18	3	1	4.2	9	32	
2	LeBronJames	\$41,180,544	6	27.13	1	2	7.3	5	6	
3	JamesHarden	\$44,310,840	2	24.92	2	3	6.7	9	7	
4	DamianLillard	\$39,344,900	9	24.63		4	6.6	5	8	
5	StephenCurry	\$45,780,966	1	24.29		5	6.5	2	9	
6	AnthonyDavis	\$35,361,360	15	23.82	2	6	2.3	4	83	
7	DevinBooker	\$31,650,600	23	23.52		7	4.6	7	28	
8	Karl-AnthonyTowns	\$31,650,600	24	23.17	,	8	3.0	8	63	
9	Kyrielrving	\$35,328,700	17	23.06	,	9	5.7	3	14	
10	RussellWestbrook	\$44,211,146	4	22.82		10	8.4	3	3	
11	BradleyBeal	\$33,724,200	18	22.06	,	12	4.1	9	35	
12	GiannisAntetokounmpo	\$39,344,900	10	21.83		13	4.6	0	29	
13	DeMarDeRozan	\$26,000,000	41	20.76	5	14	3.9	0	41	
14	PaulGeorge	\$39,344,900	7	20.39	)	15	3.5	8	50	
15	ZachLaVine	\$19,500,000	58	19.80	)	17	3.8	9	43	
16	KlayThompson	\$37,980,720	11	19.55	;	20	2.3	2	85	
17	AndrewWiggins	\$31 579 390	27	40.0		22	2.3	2	84	
- 17	Andrewwindins	.0.31 .77 3 350	27	19.26		22	/	.3	84	
owing 1	1 to 17 of 43 entries					PPGra				ank
		<b>Salary</b> \$42,018,900	SalaryRan		<b>PG</b> \$ 27.18			<b>APG</b> \$ 4.29	APGr	ank 3
owing 1	to 17 of 43 entries PlayerName	<b>Salary</b> \$42,018,900		k <sup>‡</sup> Pi	og ÷		nk ‡	APG <sup>‡</sup>		
owing 1	to 17 of 43 entries  PlayerName  KevinDurant	\$42,018,900 \$41,180,544		k • PI	<b>27.18</b>		nk <sup>‡</sup>	<b>APG</b> <sup>‡</sup> 4.29		
owing 1	PlayerName  KevinDurant LeBronJames  JamesHarden	\$42,018,900 \$41,180,544 \$44,310,840		<b>k</b> • PI 5 6 2	27.18 27.13 24.92		1 2 3	4.29 7.35 6.79		3
1 2 3 4	to 17 of 43 entries  PlayerName  KevinDurant  LeBronJames  JamesHarden  DamianLillard	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900		k * PI 5 6 2 9	27.18 27.13 24.92 24.63		1 2 3 4	4.29 7.35 6.79 6.65		3
1 2 3 4 5	PlayerName  KevinDurant LeBronJames JamesHarden DamianLillard StephenCurry	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966		k • Pi 5 6 2 9 1	27.18 27.13 24.92 24.63 24.29		1 2 3 4 5	4.29 7.35 6.79 6.65 6.52		3
1 2 3 4	to 17 of 43 entries  PlayerName  KevinDurant  LeBronJames  JamesHarden  DamianLillard	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360		k * PI 5 6 2 9	27.18 27.13 24.92 24.63		1 2 3 4	4.29 7.35 6.79 6.65		3
1 2 3 4 5 6	PlayerName  KevinDurant LeBronJames  JamesHarden DamianLillard StephenCurry AnthonyDavis	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600		k • PI 5 6 2 9 1 15	27.18 27.13 24.92 24.63 24.29 23.82		1 2 3 4 5 6	4.29 7.35 6.79 6.65 6.52 2.34		3
1 2 3 4 5 6 7	PlayerName  KevinDurant  LeBronJames  JamesHarden  DamianLillard  StephenCurry  AnthonyDavis  DevinBooker	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52		1 2 3 4 5 6 7	4.29 7.35 6.79 6.65 6.52 2.34 4.67		8 2
1 2 3 4 5 6 7 8	PlayerName  KevinDurant LeBronJames  JamesHarden DamianLillard StephenCurry AnthonyDavis DevinBooker  Karl-AnthonyTowns	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52 23.17		1 2 3 4 5 6 7 8	4.29 7.35 6.79 6.65 6.52 2.34 4.67 3.08		8 2 6
owing 1  1 2 3 4 5 6 7 8	PlayerName  KevinDurant  LeBronJames  JamesHarden  DamianLillard  StephenCurry  AnthonyDavis  DevinBooker  Karl-AnthonyTowns  Kyrielrving	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600 \$31,650,600		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52 23.17 23.06		1 2 3 4 5 6 7 8 8 9	APG 4.29 7.35 6.79 6.65 6.52 2.34 4.67 3.08 5.73		8 2 6
owing 1  2  3  4  5  6  7  8  9	PlayerName  KevinDurant  LeBronJames  JamesHarden  DamianLillard  StephenCurry  AnthonyDavis  DevinBooker  Karl-AnthonyTowns  Kyrielrving  RussellWestbrook	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600 \$31,650,600 \$35,328,700 \$44,211,146		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52 23.17 23.06 22.82		1 2 3 4 5 6 7 8 8 9 10	4.29 7.35 6.79 6.65 6.52 2.34 4.67 3.08 5.73 8.43		3 8 2 6 1
1 2 3 4 5 6 7 8 9 10 11	PlayerName  KevinDurant  LeBronJames  JamesHarden  DamianLillard  StephenCurry  AnthonyDavis  DevinBooker  Karl-AnthonyTowns  Kyrielrving  RussellWestbrook  BradleyBeal	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600 \$31,650,600 \$35,328,700 \$44,211,146		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52 23.17 23.06 22.82 22.06		1 2 3 4 5 6 7 8 9 10 12	4.29 7.35 6.79 6.65 6.52 2.34 4.67 3.08 5.73 8.43 4.19		8 2 6
owing 1 1 2 3 3 4 4 5 6 6 7 8 8 9 10 11 12	RussellWestbrook BradleyBeal GiannisAntetokounmpo	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600 \$31,650,600 \$34,211,146 \$33,724,200 \$39,344,900		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52 23.17 23.06 22.82 22.06 21.83		1 2 3 4 5 6 7 8 9 10 12 13	APG 4.29 7.35 6.79 6.65 6.52 2.34 4.67 3.08 5.73 8.43 4.19 4.60		\$\\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\
owing 1 1 2 3 4 5 6 7 8 9 10 11 12 13	PlayerName  KevinDurant LeBronJames  JamesHarden DamianLillard StephenCurry AnthonyDavis DevinBooker Karl-AnthonyTowns Kyrielrving RussellWestbrook BradleyBeal GiannisAntetokounmpo DeMarDeRozan	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600 \$31,650,600 \$31,450,600 \$31,450,600 \$31,450,600 \$31,411,146 \$33,724,200 \$39,344,900 \$39,344,900		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52 23.17 23.06 22.82 22.06 21.83 20.76		1 2 3 4 5 6 7 8 9 10 12 13 14	4.29 7.35 6.79 6.65 6.52 2.34 4.67 3.08 5.73 8.43 4.19 4.60 3.90		8 2 6 1
owing 1 1 2 3 3 4 4 5 5 6 6 7 7 8 8 9 9 10 11 12 13 14	PlayerName  KevinDurant  LeBronJames  JamesHarden  DamianLillard  StephenCurry  AnthonyDavis  DevinBooker  Karl-AnthonyTowns  Kyrielrving  RussellWestbrook  BradleyBeal  GiannisAntetokounmpo  DeMarDeRozan  PaulGeorge	\$42,018,900 \$41,180,544 \$44,310,840 \$39,344,900 \$45,780,966 \$35,361,360 \$31,650,600 \$31,650,600 \$31,450,600 \$31,450,600 \$31,450,600 \$31,450,600 \$31,450,600 \$31,44,211,146 \$33,724,200 \$39,344,900 \$26,000,000		k	27.18 27.13 24.92 24.63 24.29 23.82 23.52 23.17 23.06 22.82 22.06 21.83 20.76 20.39		1 2 3 4 5 6 7 8 9 10 12 13 14 15	4.29 7.35 6.79 6.65 6.52 2.34 4.67 3.08 5.73 8.43 4.19 4.60 3.90 3.58		3 3 4 4 5 5

5) Now combine ALL of your data into a single data frame. Your data should be in a *tidy* format. Explain what that means for your data.

### Tidy data explanation:

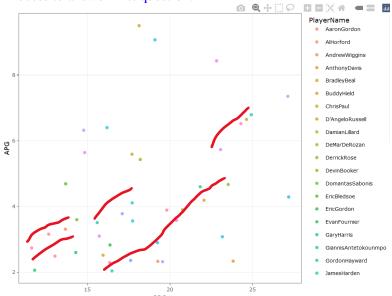
Each variable must has its own column. Each observation has its own row. Each value must has its own cell.

- 6) Choose at least three of your questions from #2 to explore. State them here. (Feel free to alter your original questions to reflect the data you were able to find and scrape.)
  - 1. PPG vs APG
  - 2. Salary vs PPG
  - 3. Who ranks in the top 100 for all three
  - 4. Salary vs APG
- 7) Now that your data is clean and tidy and ready for use, use dplyr and ggplot2 commands to extract and explore the data (numerically and graphically) in order to answer your question(s) from #6.

Explain your process as you work. What are you doing and why? In other words, what approach are you taking to answer each question (looking at what data, statistics and graphs)? Be specific. Approach your question(s) from *multiple* angles.

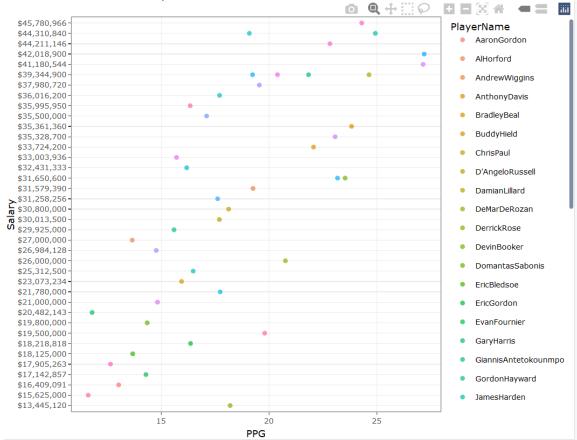
Use as many of the dplyr commands as possible, but only what makes sense for your question(s) and data.



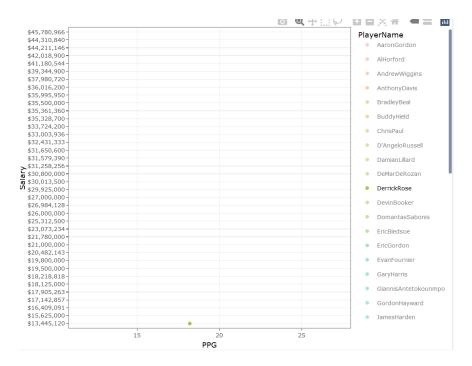


PPG vs APG. If you actually load the code then you can click on the player names on the right so it is easy to find a certain person. For this graph, I'm just comparing PPG on the x-axis and APG on the y-axis. About half of the players have under 20 PPG and below 4 APG. You can see who excels in one skill and not the other since there are some outliers such as the bottom right of about 24 PPG but not even 3 APG. For the most part, I think as one increases then so does the other. I drew some red lines above the dots to show the positive trend.

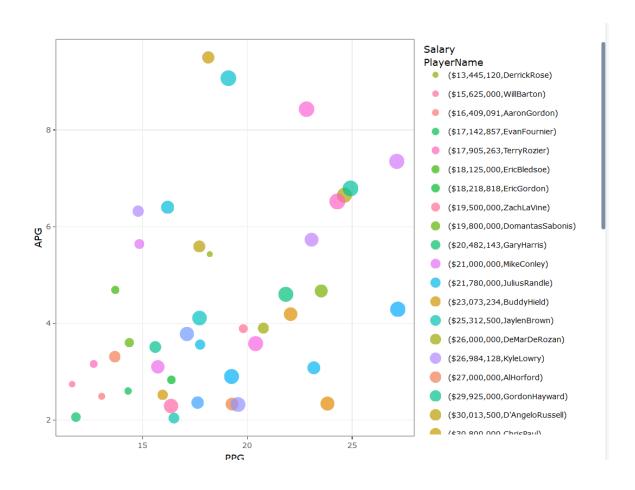
#### Process to answer second question:



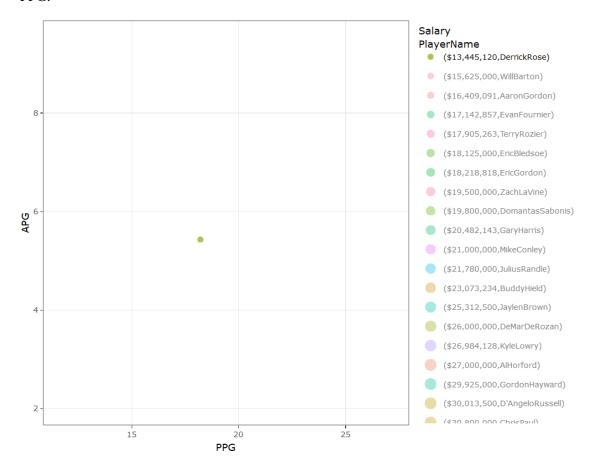
(You can click the names on the right if you load the code). This is Salary compared with PPG. This definitely has a positive relationship as almost always when PPG go up, so does the salary. You can also see who maybe should be getting paid more than others, for example: Derrick Rose has the lowest Salary out of that list of players but has more PPG than about half of the players.

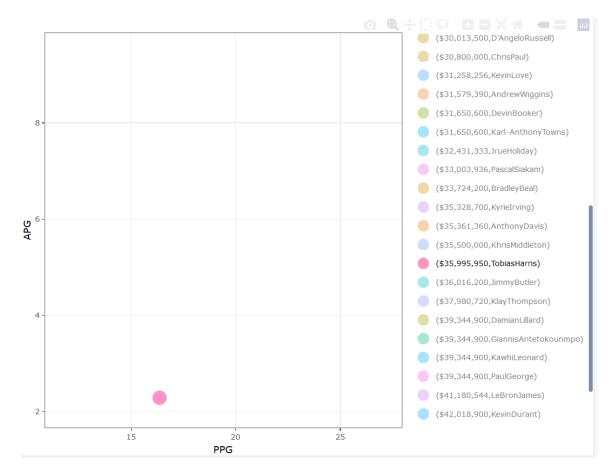


#### Process to answer third question:



(you can click on their names if you load the code). These are the 43 players who make the top 100 list in PPG, APG, and Salary rankings. Smaller bubble = smaller salary. This is easy to tell who is getting paid big bucks but not doing so well as other people who are getting paid around the same amount. Taking Derrick Rose for example again, we see that he is well above half the players for APG and leading the second half of players for PPG.





Compare Derrick Rose with Tobias Harris and we see that he is getting paid more than double what Derrick is and still has less PPG and about 3 less APG. Derrick Rose is just one person I picked out from this who should probably be earning more.

8) Now present your results! Pretend you are writing a short article on your topic/specific question(s) that will be published online. For example, see the articles from the Economist and FiveThirtyEight under in-class activities.

Your article should be 1-2 paragraphs in length, and it should include:

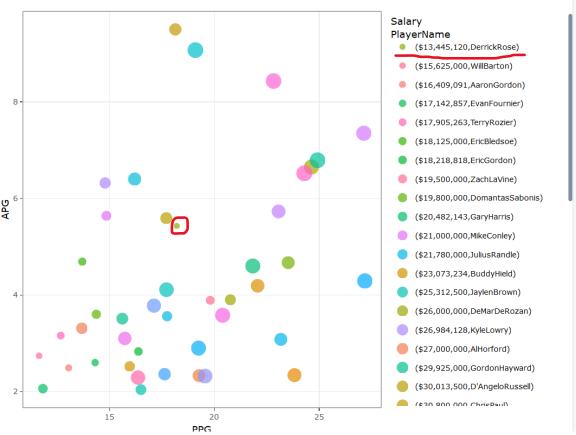
- An informative title that conveys the topic of your research.
- The answers to your questions with explanation.
- Supporting graphs, statistics, and/or other summaries. Note: You do not need to include everything from #7. Only present the most relevant information.
- A statement about where your data came from.

Bonus points will be given if your article contains a graph that displays multiple pieces of information that are clearly labeled within the graph itself. See the Joseph graph for example.

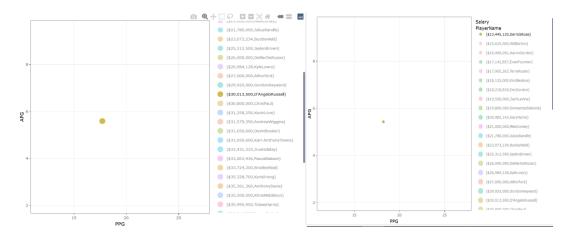
Your Article:

# Overlooked NBA players

The following data is players from the 2021-2022 NBA year who ranked within the top 100 in ALL of the following categories: PPG, APG, and yearly salary. The first players that pop into your head when thinking of the NBA are probably Steph Curry or Lebron James, and respectively so. Someone we may be overlooking though is Derrick Rose. He is the lowest paid player out of this list of 43 players who ranked in the top 100 for all three categories but he has better APG and PPG than about half of the people on the list.



Compare him with someone like D'Angelo Russell in the side-by-side comparison below and it becomes clear. They have nearly the same stats and D'Angelo is getting paid more than double. Now think about all of the bigger circles below Derrick Rose and how they are also getting paid more. One thing here stands out to me, and it is that according to these stats, Derrick Rose should be getting paid at least double what he is now for the next season.



#### Sources:

https://www.spotrac.com/nba/rankings/

https://www.basketball-reference.com/leaders/ast\_per\_g\_active.html https://www.basketball-reference.com/leaders/pts\_per\_g\_active.html