This folder contains all of the data used in The Pudding article How many high school stars make it to the NBA published in March 2019.

Below you'll find metadata for each file.

players.csv

- ```
  **What is this?**: This file contains data about the careers of all 1,873 players in our dataset.
  ```

- **Source(s) & Methods**: High school ranking data is based on the Recruiting Services Consensus Index (RSCI) rankings. All data is from Basketball Reference. We used a combination of VORP and Estimated Wins Added to determine the level of success in an NBA career, limiting to the mean ranking of each player's top five season. "Superstar" = top 30, "Great" = top 60, "Mediocre" = top 120, "Below Average" = 120+.
- **Last Modified**: March 1, 2019
- **Contact Information**: Russell Goldenberg
- **Spatial Applicability**: United States
- **Temporal Applicability**: All data was collected for players that played in a US high school between 1998 and 2013.
- **Observations (Rows)**: Each row represents data for a single player.
- **Variables (Columns)**:

Header	Description	Data Type
`name`	The player's name	text
`link`	The abbreviated link to a player's profile. If the link is written as follows: `/players/h/harrial01.html`, then that player has a profile on Basketball Reference and has made it to the NBA. Otherwise, the link is to any available player profile.	text
`rank`	Each player's high school rank. This should range between 1 and 100. Players that made it to the NBA that were not ranked in the top 100 during high school have an `NA` here.	number
`draft_year`	The year that a player declared for the NBA draft. Players that were not drafted into will have an `NA` here.	number
`draft_rd`	Denotes which round of the draft a player was selected for the NBA. Players that were not drafted will have an `NA` here.	number
`draft_pk`	Denotes which pick number overall the player was selected in the NBA draft. Players that were not drafted will have an `NA` here.	number
  ```

Header	Description	Data Type
college	The college where the player played basketball. If a player attended and played at more than one school, only the school they finished their college career in is shown. Players that did not play in college are displayed as a blank here.	text
recruit_year	The year that the player was either recruited to a college team / their last year of high school	number
bbrID	The player's Basketball Reference id number	text
pick_overall	Denotes which pick number overall the player was selected in the NBA draft. Players that were not drafted will have an NA here.	number
pick_round	Index for draft results table.	number
nba_mean_ws48	The average (mean) Win Shares Per 48 value for a player across all eligible seasons (according to required minutes played).	number
nba_mean_vorp	The average (mean) Value Over Replacement Player value for a player across all eligible seasons (according to required minutes played).	number
nba_mean_pipm	The average (mean) Player Impact Plus Minus value for a player across all eligible seasons (according to required minutes played).	number
nba_mean_wa	The average (mean) Wins Added value for a player across all eligible seasons (according to required minutes played).	number
top_mean_ws48	The average (mean) Win Shares Per 48 value for a player across their top 5 eligible seasons (according to required minutes played).	number
top_mean_vorp	The average (mean) Value Over Replacement Player value for a player across their top 5 eligible seasons (according to required minutes played).	number
top_mean_pipm	The average (mean) Player Impact Plus Minus value for a player across their top 5 eligible seasons (according to required minutes played).	number
top_mean_wa	The average (mean) Wins Added value for a player across their top 5 eligible seasons (according to required minutes played).	number
nba_mean_ws48_rank	The average (mean) league rank of Win Shares Per 48 value for a player across all eligible seasons (according to required minutes played).	number
nba_mean_vorp_rank	The average (mean) league rank of Value Over Replacement Player value for a player across all eligible seasons (according to required minutes played).	number

Header	Description	Data Type
nba_mean_pipm_rank	The average (mean) league rank of Player Impact Plus Minus value for a player across all eligible seasons (according to required minutes played).	number
nba_mean_wa_rank	The average (mean) league rank of Wins Added value for a player across all eligible seasons (according to required minutes played).	number
top_mean_ws48_rank	The average (mean) league rank of Win Shares Per 48 value for a player across their top 5 eligible seasons (according to required minutes played).	number
top_mean_vorp_rank	The average (mean) league rank of Value Over Replacement Player value for a player across their top 5 eligible seasons (according to required minutes played).	number
top_mean_pipm_rank	The average (mean) league rank of Player Impact Plus Minus value for a player across their top 5 eligible seasons (according to required minutes played).	number
top_mean_wa_rank	The average (mean) league rank of Wins Added value for a player across their top 5 eligible seasons (according to required minutes played).	
total_seasons	The total number of seasons played in the NBA.	number
valid_seasons	The number of seasons played in the NBA that meet the mininium required minutes played for advanced stats to count.	number

paths.csv & paths.R

- paths.csv: The data as described below
- paths.R: The script used to process players.csv to create paths.csv. The data was then used for the graphics in the web version of How many high school stars make it to the NBA. This script was written in the R programming language. All details describing the R session environment during processing are available here.
- ```
  **What is this?**: This file contains data about the highest level each player reached in their basketball career.
  ```
- **Source(s) & Methods**: All data is derived from players.csv.
- **Last Modified**: March 1, 2019
- **Contact Information**: Amber Thomas
- **Spatial Applicability**: United States
  ```

- **Temporal Applicability**: All data was collected for players that played in a US high school between 1998 and 2013.
- **Observations (Rows)**: Each row represents data for a single player.
- **Variables (Columns)**:

Header	Description	Data Type
`name`	The player's name	text
`college`	The college where the player played basketball. If a player attended and played at more than one school, only the school they finished their college career in is shown. Players that did not play in college are displayed as a blank here.	text
`link`	The abbreviated link to a player's profile. If the link is written as follows: `/players/h/harrial01.html`, then that player has a profile on Basketball reference and has made it to the NBA. Otherwise, the link is to any available player profile.	text
`smallMult`	Whether or not the player attended a college where at least 20 players from our dataset attended. `0` = they did not, `1` = they did.	number
`rank`	Each player's high school rank. This should range between 1 and 100. Players that made it to the NBA that were not ranked in the top 100 during high school have an `NA` here.	number
`draft_year`	The year that a player was drafted into the NBA. Players that were not drafted into the NBA, will have an `NA` here.	number
`draft_rd`	Denotes which round of the draft a player was selected for the NBA. Players that were not drafted into the NBA will have an `NA` here.	number
`draft_pk`	Denotes which pick number in the round of the draft (as indicated by `draft_rd`) of the draft a player was selected for the NBA. Players that were not drafted into the NBA will have an `NA` here.	number
`recruit_year`	The year that the player was either recruited to a college team or their last year of high school or prep school	number
`avgRank`	The average (mean) of the `top_mean_vorp_rank` and `top_mean_wa_rank` for each player	number
`top`	Whether the player was ranked in the top 100 when they were in high school (`0` = no, `1` = yes)	number
`highSchool`	Whether or not a player continued to play basketball beyond their high school career (`0` = no, `1` = yes)	number

Header	Description	Data Type
coll	Whether or not a player continued to play basketball beyond their college career (0 = no, 1 = yes, 2 = skipped college and went straight to NBA from high school, NA = didn't make it to college level)	number
draft	Whether or not a player continued to play basketball beyond being drafted (0 = no, 1 = yes, NA = didn't make it to draft level)	number
rookie	Whether or not a player continued to play basketball the first 2 years in the NBA (0 = no, 1 = yes, NA = didn't make it to NBA level)	number
valid_seasons		number
allstar	Whether or not a player reached allstar status in the NBA (avgRank < 30) (0 = no, 1 = yes, NA = didn't make it past rookie level)	number
bad	Whether or not a player reached "below average" status in the NBA (avgRank > 120) (0 = no, 1 = yes, NA = didn't make it past rookie level)	number
good	Whether or not a player reached good status in the NBA (avgRank < 120 && avgRank > 60) (0 = no, 1 = yes, NA = didn't make it past rookie level)	number
great	Whether or not a player reached great status in the NBA (avgRank < 60 && avgRank > 30) (0 = no, 1 = yes, NA = didn't make it past rookie level)	number
highest	The string value for the highest level reached (highSchool, college, draft, rookie, bad, good, great, allstar)	text

Session info ----------

```
##  setting  value
##  version  R version 3.4.4 (2018-03-15)
##  system   x86_64, darwin15.6.0
##  ui       X11
##  language (EN)
##  collate  en_US.UTF-8
##  tz       America/Los_Angeles
##  date     2019-03-04

## Packages ---------------------------------------------------------------

##  package    * version   date       source
##  assertthat   0.2.0     2017-04-11 CRAN (R 3.4.0)
##  backports    1.1.2     2017-12-13 CRAN (R 3.4.3)
##  base       * 3.4.4     2018-03-15 local
##  bindr        0.1.1     2018-03-13 CRAN (R 3.4.4)
##  bindrcpp     0.2.2     2018-03-29 CRAN (R 3.4.4)
##  broom        0.4.4     2018-03-29 CRAN (R 3.4.4)
##  cellranger   1.1.0     2016-07-27 CRAN (R 3.4.0)
```

```
##  cli          1.0.0     2017-11-05 CRAN (R 3.4.2)
##  colorspace   1.3-2     2016-12-14 CRAN (R 3.4.0)
##  compiler     3.4.4     2018-03-15 local
##  crayon       1.3.4     2017-09-16 CRAN (R 3.4.1)
##  datasets   * 3.4.4     2018-03-15 local
##  devtools     1.13.6    2018-06-27 CRAN (R 3.4.4)
##  digest       0.6.15    2018-01-28 CRAN (R 3.4.3)
##  dplyr      * 0.7.8     2018-11-10 cran (@0.7.8)
##  evaluate     0.10.1    2017-06-24 CRAN (R 3.4.1)
##  forcats    * 0.3.0     2018-02-19 CRAN (R 3.4.3)
##  foreign      0.8-69    2017-06-22 CRAN (R 3.4.4)
##  ggplot2    * 3.1.0     2018-10-25 CRAN (R 3.4.4)
##  glue         1.3.0     2018-07-17 CRAN (R 3.4.4)
##  graphics   * 3.4.4     2018-03-15 local
##  grDevices  * 3.4.4     2018-03-15 local
##  grid         3.4.4     2018-03-15 local
##  gtable       0.2.0     2016-02-26 CRAN (R 3.4.0)
##  haven        1.1.1     2018-01-18 CRAN (R 3.4.3)
##  here       * 0.1       2017-05-28 CRAN (R 3.4.0)
##  hms          0.4.2     2018-03-10 CRAN (R 3.4.4)
##  htmltools    0.3.6     2017-04-28 CRAN (R 3.4.0)
##  httr         1.3.1     2017-08-20 CRAN (R 3.4.1)
##  jsonlite     1.6       2018-12-07 CRAN (R 3.4.4)
##  knitr        1.20      2018-02-20 CRAN (R 3.4.3)
##  lattice      0.20-35   2017-03-25 CRAN (R 3.4.4)
##  lazyeval     0.2.1     2017-10-29 CRAN (R 3.4.2)
##  lubridate    1.7.4     2018-04-11 CRAN (R 3.4.4)
##  magrittr     1.5       2014-11-22 CRAN (R 3.4.0)
##  memoise      1.1.0     2017-04-21 CRAN (R 3.4.0)
##  methods    * 3.4.4     2018-03-15 local
##  mnormt       1.5-5     2016-10-15 CRAN (R 3.4.0)
##  modelr       0.1.1     2017-07-24 CRAN (R 3.4.1)
##  munsell      0.5.0     2018-06-12 cran (@0.5.0)
##  nlme         3.1-131.1 2018-02-16 CRAN (R 3.4.4)
##  parallel     3.4.4     2018-03-15 local
##  pillar       1.2.2     2018-04-26 CRAN (R 3.4.4)
##  pkgconfig    2.0.2     2018-08-16 cran (@2.0.2)
##  plyr         1.8.4     2016-06-08 CRAN (R 3.4.0)
##  psych        1.8.3.3   2018-03-30 CRAN (R 3.4.4)
##  purrr      * 0.2.5     2018-05-29 cran (@0.2.5)
##  R6           2.3.0     2018-10-04 cran (@2.3.0)
##  Rcpp         1.0.0     2018-11-07 cran (@1.0.0)
##  readr      * 1.1.1     2017-05-16 CRAN (R 3.4.0)
##  readxl       1.1.0     2018-04-20 CRAN (R 3.4.4)
##  reshape2     1.4.3     2017-12-11 CRAN (R 3.4.3)
##  rlang        0.3.0.1   2018-10-25 cran (@0.3.0.1)
##  rmarkdown    1.9       2018-03-01 CRAN (R 3.4.3)
##  rprojroot    1.3-2     2018-01-03 CRAN (R 3.4.3)
##  rstudioapi   0.7       2017-09-07 CRAN (R 3.4.1)
##  rvest        0.3.2     2016-06-17 CRAN (R 3.4.0)
##  scales       1.0.0     2018-08-09 cran (@1.0.0)
```

```
##  stats      * 3.4.4     2018-03-15 local
##  stringi      1.1.7     2018-03-12 CRAN (R 3.4.4)
##  stringr    * 1.3.1     2018-05-10 CRAN (R 3.4.4)
##  tibble     * 1.4.2     2018-01-22 CRAN (R 3.4.3)
##  tidyr      * 0.8.0     2018-01-29 CRAN (R 3.4.3)
##  tidyselect   0.2.5     2018-10-11 cran (@0.2.5)
##  tidyverse  * 1.2.1     2017-11-14 CRAN (R 3.4.2)
##  tools        3.4.4     2018-03-15 local
##  utils      * 3.4.4     2018-03-15 local
##  withr        2.1.2     2018-03-15 CRAN (R 3.4.4)
##  xml2         1.2.0     2018-01-24 CRAN (R 3.4.3)
##  yaml         2.1.18    2018-03-08 CRAN (R 3.4.4)
```