

MAT 259: Project 3 Codebook

Brianna Griffin

Below is what the first few rows of the data set that I am using for Project 3 look likes. It contains metrics on MLB pitchers in the National League from 2000-2022.

```
##      age  tm      ip nd wchp ltuf wtm ltm tm_w_l_percent wlst lsv cg sho qs
## 1 25-29 ARI 213.1 14    2    2 19 13          0.594    7  4  2  0 20
## 2 19-24 PIT 144.0 10    1    3  9 17          0.346    3  5  1  0 10
## 3 19-24 STL 175.0 12    2    1 18 12          0.600    3  5  0  0 16
## 4 19-24 MON  95.0  1    2    1  8  9          0.471    0  0  0  0  7
## 5 25-29 CHC  32.2  1    0    0  1  3          0.250    1  0  0  0  0
## 6 30-34 COL 101.1  5    1    2  9 10          0.474    1  3  0  0  7
##      qs_percent gm_sc_a best wrst bqr s_dr l_dr rs_gs ip_gs pit_gs x80 x80_99
## 1          63    50.1   78   12 22    0 13   5.1   6.5    93   5    18
## 2          38    44.2   78   20 19    1 17   4.7   5.4    87   6    14
## 3          53    55.3   75   15  8    1 16   5.3   5.8   101   1    13
## 4          41    50.5   81   23 19    0  9   4.6   5.6    88   5     8
## 5           0    32.8   45   23 12    1  2   3.2   3.8    73   2     2
## 6          37    42.9   64   18  9    1  9   5.6   5.3    94   4     6
##      x100_119 x120 max year cy_status      w_pct      arm
## 1           7    2 130 2000          no 0.3333333 left
## 2           3    3 128 2000          no 0.1851852 left
## 3          15    1 120 2000          no 0.3548387 left
## 4           4    0 112 2000          no 0.4117647 right
## 5           0    0  92 2000          no 0.0000000 right
## 6           9    0 119 2000          no 0.2631579 right
```

Now, I would like to explain the variables that I am utilizing in my Project 3 visualization for clarity.

- **Winning Percentage** = number of games won ÷ total number of games pitched
- **Innings Pitched** = numbers of innings pitched during the season
- **Year** = year of the given season
- **Best** = best game score
- **Worst** = worst game score
 - **Game Score** measures a pitcher's performance in any given game started
- **WLST** = number of wins lost
 - At the time the pitcher faced his final batter the pitcher was in position for a win, but game was blown by bullpen.
- **LSV** = number of losses saved
 - At the time of his last batter the pitcher was in position for a loss, but team came back to tie or take lead.
- **Team Win-Loss %** = team win loss percentage
 - The win-loss percentage of the team in games started by this pitcher.
- **Quality Start %** = quality start percentage
 - Percentage of starts that were quality starts: pitcher pitched at least 6 innings and allowed 3 or fewer earned runs in a start.
- **# Short Days Rest** = number of short days rest

- less than 4 days of rest
- **# Long Days Rest** = number of long days rest
 - more than 4 days of rest

Hopefully this helps clear up any confusion with the data and visualization.