

3. Understanding Your Data

The table below contains a data set that describes the top 25 salaries for major league baseball players as of opening day of the 2016 season:

<i>Player</i>	<i>Team</i>	<i>Position</i>	<i>Age (as of April 28, 2016)</i>	<i>Salary</i>
Kershaw	Dodgers	Pitcher	28	34.57
Greinke	Diamond Backs	Pitcher	32	34
Price	Red Sox	Pitcher	30	30
Verlander	Tigers	Pitcher	33	28
Cabrera	Tigers	1st Base	33	28
Hernandez	Mariners	Pitcher	30	25.85
Sabathia	Yankees	Pitcher	35	25
Lester	Cubs	Pitcher	32	25
Howard	Phillies	1st base	36	25
Pujols	Angels	Designated Hitter	36	25
Cano	Mariners	2nd Base	33	24
Hamels	Rangers	Pitcher	32	23.5
Teixeira	Yankees	1st Base	36	23.13
Mauer	Twins	1st Base	33	23
Ramirez	Red Sox	1st Base	32	22.75
Scherzer	Nationals	Pitcher	31	22.14
Upton	Tigers	Left Field	28	22.13
Tanaka	Yankees	Pitcher	27	22
Reyes	Rockies	Shortstop	32	22
Gonzalez	Dodgers	1st Base	33	21.86
Crawford	Dodgers	Left Field	34	21.61
Werth	Nationals	Left Field	36	21.57
Ellsbury	Yankees	Center Field	32	21.14
Davis	Orioles	1st Base	30	21.12
Shields	Padres	Pitcher	34	21

- (1) What are the individuals (i.e., items of analysis) in this data set? Consider what each row in the table above represents to answer this question.
- (2) In addition to the players' names, how many variables does the data set contain? Which of these variables take numerical values?
- (3) What are the units in which each of the numerical values is expressed? For example, what does it mean when Howard's salary is listed as 25?
- (4) What is the most common position in the data set? What is the most common salary? Do you think the most common salary will be the same as the average salary? Why or why not?