Lab 03 - Merging Data

Francisco Santamarina September 10, 2016

Load the necessary packages and dataset.

```
##load data
library(Lahman)
```

Question 1

What is the average salary of a player on the Oakland A's in 2002?

\$1,482,000

A player on the Boston Red Sox that same year?

\$3,612,000

What is the total budget of each franchise in 2002? Oakland A's: \$40,004,167 Boston Red Sox: \$108,366,060

To answer the question, merge the Salaries and Teams datasets.

sum(dat.salaries.teams\$salary [OAK2002.salaries])

```
data( Master )
data( Teams )
data( Salaries )
dat.salaries.teams <- merge( Salaries, Teams )</pre>
OAK2002.salaries <- dat.salaries.teams$name == "Oakland Athletics" & dat.salaries.teams$yearID == "2002
#mean( dat.salaries.teams$salary [OAK2002.salaries], na.rm=T )#produces weird values
summary(dat.salaries.teams$salary [OAK2002.salaries])
##
      Min. 1st Qu.
                   Median
                              Mean 3rd Qu.
   202000 398800 812500 1482000 1662000 7167000
#find players on Red Sox in 2002
#merge that with salary
#find average
BOS2002.salaries <- dat.salaries.teams$name == "Boston Red Sox" & dat.salaries.teams$yearID == "2002"
#mean( dat.salaries.teams$salary [BOS2002.salaries], na.rm=T )#produces weird values
summary(dat.salaries.teams$salary [BOS2002.salaries])
##
                       Median
                                  Mean 3rd Qu.
                                                    Max.
       Min.
            1st Qu.
              412500 2288000 3612000 5750000 15460000
     200000
#find the franchises in 2002
#find the salaries for each one
#sum the salaries for each one
#print as a table
```

```
## [1] 40004167
```

```
sum( dat.salaries.teams$salary [BOS2002.salaries] )
## [1] 108366060
```

Question 2

2. Which variables are used for the default merge? Why would the following merge be problematic?

```
merge( Teams, Salaries, by.x="teamID", by.y="teamID")
```

The variables being used are teamID from Teams and teamID from Salaries. * This uses Teams as the primary table. Any independent players whose salaries are not associated to a team, or with a teamID, will not appear in the merged dataset. * This assumes that teamID are the same uniqueID and are the appropriate variables to use * This merge will not include teams without salaries.

Question 3

3. Which city (birth city) has produced the most baseball hall-of-famers? Use the Master and HallOfFame datasets.

San Francisco, with 128 players

To check which variables will be used for the default merge: intersect(names(HallOfFame), names(Master))

```
library(Lahman)
data( Master )
data( HallOfFame )
intersect( names(HallOfFame), names(Master) )
```

```
## [1] "playerID"
```

```
ls(Master)
```

```
[1] "bats"
                        "bbrefID"
                                        "birthCity"
                                                        "birthCountry"
    [5] "birthDate"
                        "birthDay"
                                        "birthMonth"
                                                        "birthState"
    [9] "birthYear"
                        "deathCity"
                                        "deathCountry"
                                                        "deathDate"
## [13]
        "deathDay"
                        "deathMonth"
                                        "deathState"
                                                        "deathYear"
       "debut"
                        "finalGame"
                                        "height"
                                                        "nameFirst"
## [17]
  [21] "nameGiven"
                        "nameLast"
                                        "playerID"
                                                        "retroID"
## [25] "throws"
                        "weight"
```

```
master.halloffame <- merge( Master, HallOfFame, by.x = "playerID", by.y = "playerID")
sort( table( master.halloffame$birthCity))</pre>
```

```
## ## Aberdeen Abington Adrian ## 1 1 1
```

##	Aguada	Aguadilla	Alameda
##	1 ^7 di -	1	1
## ##	Alexandria 1	Alger	Aliquippa 1
##	Amherst	Antioch	Appleton
##	1	1	1
##	Arcadia	Arkansas City	Arlington
##	1	1	1
##	Arno	Arroyo	Atkins
##	1	1	1
##	Avon Park	Bakersfield	Ballston Spa
##	1	1	1
##	Bancroft	Bani	Barceloneta
##	D-+	1	D
## ##	Batesville 1	Bay 1	Beaver Dam 1
##	Bellaire	Belleville	Bishop
##	1	1	1
##	Bloomingdale	Bonao	Boynton Beach
##	1	1	1
##	Bremerton	Brenham	Bridgeport
##	1	1	1
##	Brighton	Brooksville	Brunswick
##	1	1	1
##	Buena Vista	Burlingame	Burlington Flats
##	1	1	<u></u>
## ##	Cairo 1	Canaan 1	Cardenas 1
##	Caripito	Carolina	Casper
##	1	1	oasper 1
##	Cass City	Catano	Cecil County
##	1	1	1
##	Central Alava	Centralia	Charlestown
##	1	1	1
##	Chattanooga	Chestertown	Chicago Heights
##	1	1	1
##	Chico	Childersburg	Chouteau
##	1 Ch	1 C:f	1 (1) - i - i - i - i - i - i - i - i - i -
## ##	Churuguara 1	Cienfuegos 1	Clairton 1
##	Clarksboro	Clearbrook	Clearwater
##	1	1	1
##	Cochecton	Coffeyville	Cohoes
##	1	1	1
##	Connellsville	Cordova	Corydon
##	1	1	1
##	Council Bluffs	Covina	Cushing
##	1	1	1
##	Dallas	Denton	Denver
##	Des Disines	1 Di	1 Di
##	Des Plaines	Dixon 1	Dorris 1
##	Dubuque	Eagle Pass	Earlsboro
##	Dubuque 1	Lagie rass	1
	-	-	-

## ##	East Chicago	East Haddam 1	East Liverpool 1
##	East Providence	Eau Claire	Edgewater
##	1	1	1
##	Ehime	El Dorado	El Seibo
##	1	1	1
##	Elgin 1	Ellisville	Elmwood Place 1
## ##	Emlenton	1 Etna	Evergreen
##	1	1	Lvergreen 1
##	Ewing	Exeter	Factoryville
##	1	1	1
##	Fall River	Fayetteville	Flat Rock
##	1	1	1
##	Flint	Floral Park	Florence
##	1	1	1
## ##	Fontana 1	Forrest City	Fort Gibson 1
##	Fort Pierce	Fort Rucker	Fort Worth
##	1	1	1
##	Fostoria	Franklinton	Fredericksburg
##	1	1	1
##	Fredericton	Freiburg	Fulton
##	1	1	1
##	Galena	Galesburg	Galveston
## ##	1 Carr	1 Gatun	Causa
##	Gary 1	Gatun 1	Gause 1
##	Giddings	Glen Dale	Golden Valley
##	1	1	1
##	Goshen	Granada	Granard
##	1	1	1
##	Grand Rapids	Granite City	Greenfield
##	1	1	1
##	Gulfport	Haleburg	Hall Summit
## ##	1 Hamtramck	1 Hanford	1 Hanover
##	nameramek 1	1	nanover 1
##	Hatboro	Haverhill	Havre de Grace
##	1	1	1
##	Hazleton	Hempstead	Henning
##	1	1	1
##	Herlong	Hickory Grove	Hillsboro
##	1	1 1	1
## ##	Holguin 1	Holly Hill 1	Hollywood 1
##	Homestead	Hoopeston	Hoyleton
##	nomestead 1	noopes con	noyleton 1
##	Humboldt	Huntington	Huntington Beach
##	1	1	1
##	Independence	Inglewood	Jacksonville
##	1	1	1
##	Johnstown	Joliet	Joplin
##	1	1	1

шш	I	T	17
## ##	Junction City	Justiceburg 1	Kearny 1
##	Kensington	Key West	Keytesville
##	Rensington 1	ney west	Reytesville
##	Killorgin	La Mesa	Lacoochee
##	1	1	1
##	Lakeland	Las Villas	Lebanon
##	1	1	1
##	Leipsic	Lenoir	Lexington
##	1	1	1
##	Little Rock	Livermore	Lockney
##	1	1	1
##	Lompoc	Longview	Los Molinos
##	1	1	1
##	Lowell	Magnolia	Manoguayabo
##	1	1	1
##	Marion	Marlborough	Marshall
##	1	1	1
##	Martins Ferry	Martinsville	Maryland Heights
##	1	1	1
##	McKeesport	Media	Melville
##	Ma+1	Minut Desch	1 W: 43 4
##	Methuen	Miami Beach	Midland
## ##	Milford	M:::::::::::::	Minneanelia
##	Milford 1	Millville 1	Minneapolis 1
##	Modesto	Monte Cristi	Montgomery
##	1	1	1
##	Monticello	Montreal	Montrose
##	1	1	1
##	Montville	Moosup	Morgan City
##	1	1	1
##	Mount Morris	Muskegon	Narrows
##	1	1	1
##	New Athens	New Brighton	New Brunswick
##	1	1	1
##	New Castle	New Ulm	Newton
##	1	1	1
##	Ney	Niles	Norwood
##	1	1	1
##	Oak Grove	Daxaca	Ocumare del Tuy
##	1	1	1
##	Olyphant	Opelika	Osaka
##	1 Oxford	1 Oxnard	Dolo+ko
## ##		Uxnard 1	Palatka 1
##	Panola	Pascagoula	Pasco
##	7 anoia 1	rascagoula 1	rasco 1
##	Passaic	Pelican Rapids	Pendleton
##	1	1 erican napius	1 endrecon
##	Pensacola	Perry	Perryton
##	1	1	1
##	Perryville	Phoenix	Pittsburg
##	1	1	1

##	Plainfield	Plateau	Pontiac
##	1	1	1
##	Port Huron	Prescott	Providence
## ##	1 Pueblo Nuevo	1 Pueblo Viejo	1 Qualls
##	Pueblo Nuevo	ruebio viejo	Qualis 1
##	Quincy	Ramer	Ramon Santana
##	1	1	1
##	Refugio	Remington	Ridgeway
##	1	1	1
##	Rio Piedras	Rock Hill	Rocky Mount
##	1	1	1
##	Rowes Run	Russellton	Rybnik
##	1	1	1
##	Salinas	Salt Lake City	San Angelo
## ##	1 San Fernando	1 San Gabriel	1 San Mateo
##	Jan Fernando	San Gabrier	Jan Hateo
##	San Pedro	Sandcut	Sandy Lake
##	1	1	1
##	Santa Barbara	Santa Paula	Schenectady
##	1	1	1
##	Sebring	Seguin	Seneca Falls
##	1	1	1
##	Shinnston	Silver Spring	Sissonville
##	1 C H:11	1	1 Cauthamatan
## ##	Snow Hill 1	South Gate	Southampton 1
##	Spavinaw	St. Cloud	St. George
##	5pavinaw 1	1	1
##	Staten Island	Staunton	Stockton
##	1	1	1
##	Story City	Sunnyside	Swansea
##	1	1	1
##	Sydney	Tarzana	Tiburon
##	1	1	1
##	Tiskilwa	Toronto	Torrance
## ##	1 Tulare	1 Tulsa	1 Tuolumne
##	1	1 ursa	
##	Tuskegee	Tuxedo	Union City
##	1	1	1
##	Uniontown	Vallejo	Van Meter
##	1	1	1
##	Vancouver	Venice	Vicksburg
##	1	1	1
##	Vienna	Vinita	Waterbury
##	1	1	1
## ##	Wawaka 1	Wellston 1	West Bradford 1
##	West Chester	Westfield	Willingboro
##	west chester	westileid 1	willingboro 1
##	Winter Haven	Wolf Point	Wyandotte
##	1	1	1

##	Yabucoa	Albany	Arnot
##	1	2	2
##	Bedford	Bellefonte	Bellows Falls
##	2	2	2
##	Bethpage 2	Boyceville 2	Breckenridge 2
## ##	Byron	Calvert	Caney Springs
##	2 by 1011	2	caney springs
##	Caroleen	Carrick	Cassville
##	2	2	2
##	Charleston	Charlotte	Chartiers
##	2	2	2
##	Clanton	Clinton	Concord
##	2 Covington	2 Cmag+line	2 Engt Prophiald
## ##	Covington 2	Crestline 2	East Brookfield 2
##	East St. Louis	Egypt	El Cajon
##	2	2	2
##	Euclid	Frederiksted	French Creek
##	2	2	2
##	Graham	Grants Pass	Hackensack
##	2	2	2
##	Hazel Park	Helena	Highland Park
## ##	2 Hollis	2 Hubbard	2 Jackson
##	2	nubbard 2	Jackson 2
##	Jefferson City	Kingston	Long Island City
##	2	2	2
##	Los Gatos	Louisville	Lynn
##	2	2	2
##	Lynwood	Macon	Marshalltown
##	2	2	2
## ##	Massena 2	Merced 2	Meriden 2
##	Milledgeville	Mineral	Monroe
##	2	2	2
##	Mount Vernon	Navojoa	Nelliston
##	2	2	2
##	Newport	Niagara Falls	Norwalk
##	2	2	2
##	Omaha	Orange	Orlando
##	Do go dono	2 Dataman	Dhillinghung
## ##	Pasadena 2	Paterson 2	Phillipsburg 2
##	Pickens County	Pine City	Pryor
##	2	2	2
##	Ragland	San Antonio	Savannah
##	2	2	2
##	Sheffield	Sheridan	St. Charles
##	2	2	2
##	St. Mary	Starkville	Steubenville
##	2 Sulvan Laka	2 Ta.coma	Takama Bark
## ##	Sylvan Lake 2	Tacoma 2	Takoma Park 2
##	2	2	2

##	Tarboro	Topeka	Victoria
##	2	2	2
##	Vinton	Walnut Creek	Ware
##	2	2	2
##	West Palm Beach	Westminster	Whittier
##	2	2	2
##	Youngstown	Zerbe	Austin
##	2	2	3
##	Bajos de Haina 3	Beaumont 3	Burlington
## ##	Caracas	Caseyville	3 Chatham
##	3	Caseyville 3	3
##	Colon	Danville	Donora
##	3	3	3
##	Dublin	Elba	Flat
##	3	3	3
##	Gilmore	Hertford	Hoagland
##	3	3	3
##	Honolulu	Laguna Verde	Marietta
##	3 Meadville	3 New Bethlehem	Nambh Darahfiald
## ##	MeadVIIIe	New Bethlenem	North Brookfield 3
##	Pittsfield	Richmond	San Cristobal
##	3	3	3
##	Santa Maria	Santo Domingo	Santurce
##	3	3	3
##	Smithtown	Spokane	Sumter
##	3	3	3
##	Toms River	Upper Chanceford	Waukegan
##	3	3	3
##	West Lebanon	Williamsport	Winston-Salem
## ##	3 Witherbee	3 Woodsfield	Anahamama
##	witherbee 3	woodsileid 3	Anchorage 4
##	Arecibo	Bellflower	Berwyn
##	4	4	4
##	Billings	Binghamton	Burbank
##	4	4	4
##	Carthage	Decatur	Du Bois
##	4	4	4
##	Durham	Fitchburg	Fullerton
##	4 C	4 C	4
## ##	Greenville 4	Gretna 4	Hannibal 4
##	Mansfield	Matanzas	Millerton
##	4	4	4
##	Montclair	Newark	Ontario
##	4	4	4
##	Pageland	Payette	Peoria
##	4	4	4
##	Portsmouth	Red Oak	Rural Retreat
##	4	4	4
##	San Juan	Santa Ana	Scranton
##	4	4	4

##	Stefurov	Truxton	Tuscaloosa
##	4	4	4
## ##	Villa Rica 4	Waynesboro 4	Beach Haven 5
##	Blaine	Bonham	Caton
##	5	5	5
##	Clio	Columbia	Commerce
##	5	5	5
##	Easton	Franklin	Georgetown
##	5	5	5
##	Hartford	Lawrence	Le Grand
##	5	5	5
##	Lewiston	Long Beach	Martinez
##	5	5	5
##	Mount Ida	New Haven	Norristown
##	Nameth Adams	Ohlahama Cita	5
## ##	North Adams 5	Oklahoma City 5	Pomeroy 5
##	Pottstown	Poway	Santa Monica
##	5	5 - 5	5
##	Shinglehouse	Shreveport	South Wales
##	5	5	5
##	Stony Creek Mills	Streetman	Texarkana
##	5	5	5
##	Trenton	Waco	Williamston
##	5	5	5
##	Wilmington	Winters	Ash Flat
##	5	5	6
##	Birmingham	Cedar Rapids 6	Culver City 6
## ##	6 Greensboro	Lonaconing	Malden
##	Greensboro 6	Lonaconing 6	Marden 6
##	Manchester	Maple Ridge	Maracaibo
##	6	6	6
##	Mobile	Nashville	Syracuse
##	6	6	6
##	Uvalda	Whistler	Auburn
##	6	6	7
##	Bell Station	Bolton	Broken Bow
##	7	7	7
##	Crabapple	Cranston	Gettysburg
## ##	7 Gordonsville	7 Hinckley	7 Linden
##	GOLGORSVILLE 7	ninckiey 7	Linden 7
##	Meridian	Minooka	Mishawaka
##	7	7	7
##	Mooresville	Pittston	Toledo
##	7	7	7
##	Tuscumbia	Bridgewater	Camden
##	7	8	8
##	Clarendon	Clear Creek	Collinsville
##	8	8	8
##	Cornwall	Dayton	Elkhart
##	8	8	8

##	Fowlerville	Gloucester	High Point
##	8	8	8
##	Huntersville	Muscotah	New Orleans
##	8	8	8
##	Penacook 8	Sudlersville 8	Utica 8
## ##	Vernon	Wahoo	Winterset
##	vernon 8	wanoo 8	winterset 8
##	Woodland	Bearsville	Boston
##	8	9	9
##	Bradford	Camaguey	Colorado Springs
##	9	9	9
##	Framingham	Harvard	Kennett Square
##	9	9	9
##	Lafayette	Mahanoy City	McSherrystown
##	9 N	9	9
## ##	Nyesville 9	Plains 9	Silver Creek 9
##	Snohomish	Wooster	Havana
##	9	9	10
##	Huntsville	Massillon	Medway
##	10	10	10
##	New Albany	Pomona	Salem
##	10	10	10
##	San Pedro de Macoris	Sanford	York
##	10	10	10
##	Bronx 11	Carteret 11	Lake Charles
## ##	Lucas	Port Jervis	Troy
##	11	roit Jeivis	1109
##	Wilkinsburg	Archie	Bastrop
##	11	12	12
##	Harvey	Miami	Sacramento
##	12	12	12
##	Shamokin	Titus	Van Nuys
##	12	12	12
##	Baltimore	Beemer	Clayton
## ##	13	13	13
##	Harrisville 13	Hartsville 13	Oglesby 13
##	San Jose	Trappe	Waldo
##	13	13	13
##	Atherton	Bluffton	Cambridge
##	14	14	14
##	Clifty	Demorest	Lancaster
##	14	14	14
##	Morocco	Prospect Park	Richburg
##	14	14	14
##	San Diego	Wadena	Wampum
## ##	14 Zeist	14 Argo	14 Bethany
##	Zeist 14	Argo 15	bethany 15
##	Emerald	Fresno	Glasgow
##	15	15	15

##	Grenada	Hibbing	Marianao
##	15	15	15
##	Michigan City	Milwaukee	Nitro
##	15	15	15
##	Ocumare de la Costa	Pebworth	Princeton
##	Dochaster	15	Santa Bita
## ##	Rochester 15	San Bernardino 15	Santa Rita 15
##	Stephentown	Temple	Tipton
##	15	15	15
##	West Allis	Zeeland	Berkeley
##	15	15	16
##	Comanche	Crow Wing County	Evansville
##	16	16	16
##	Garden Grove	Germantown	Jamestown
##	16	16	16
##	Liberty	Marcus Hook	Pinar del Rio
##	16	16	16
##	Riverside	Rodeo	Roxboro
##	16	16	16
##	Sioux City	Southgate	St. Thomas
##	16	16	16
##	Swifton	Tilden	Woonsocket
##	16	16	16
## ##	Cascade 17	Ekron 17	Ellwood City
##	Hamilton	Middletown	Orient
##	17	17	17
##	Pittsburgh	St. Paul	Culpeper
##	17	17	18
##	Granville	Harrah	Harvel
##	18	18	18
##	Houston	Memphis	Seattle
##	18	18	19
##	Shiner	Washington	West Springfield
##	19	19	19
##	Akron	Oakland City	Ponce
##	20	20	20
##	Columbus	Indianapolis	Atlanta
##	21	21	22
##	Anderson	Wheeling	Madison
## ##	24 Buffalo	24	25 Cincinnati
##	26	Detroit 29	32
##	Oakland	Portland	Springfield
##	32	32	32
##	Kansas City	La Habana	Terre Haute
##	33	37	41
##	Tampa	Cleveland	New York
##	42	45	58
##	Los Angeles	Philadelphia	St. Louis
##	63	69	72
##	Chicago	Brooklyn	San Francisco
##	74	79	128