

Assignment 5

edelsonc

September 8, 2016

This assignment will look at the aggregate function. We will look at this while investigating the difference between the salaries of female and male professor at an unnamed university. First the file is downloaded from the web.

```
download.file("https://vincentarelbundock.github.io/Rdatasets/csv/car/Salaries.csv",
              "/Users/edelsonc/Desktop/Data_Science/EDA/assignment_5/Salaries.csv")
```

Then it is read into memory using `read.csv`

```
f_path <- "/Users/edelsonc/Desktop/Data_Science/EDA/assignment_5/Salaries.csv"

Salaries <- read.csv(f_path, header=TRUE)

str(Salaries)
```

```
## 'data.frame':   397 obs. of  7 variables:
## $ X             : int  1 2 3 4 5 6 7 8 9 10 ...
## $ rank           : Factor w/ 3 levels "AssocProf","AsstProf",...: 3 3 2 3 3 1 3 3 3 3 ...
## $ discipline     : Factor w/ 2 levels "A","B": 2 2 2 2 2 2 2 2 2 2 ...
## $ yrs.since.phd  : int  19 20 4 45 40 6 30 45 21 18 ...
## $ yrs.service    : int  18 16 3 39 41 6 23 45 20 18 ...
## $ sex            : Factor w/ 2 levels "Female","Male": 2 2 2 2 2 2 2 2 2 1 ...
## $ salary         : int  139750 173200 79750 115000 141500 97000 175000 147765 119250 129000 ...
```

Dependence of Salary on Other Variables

We can begin to investigate the effects of rank on salary using the `aggregate` function in R. Aggregating data is the simplest form of analysis, and using it we can look at how different groups of sex and rank receive different salaries

```
aggregate(salary ~ rank + sex, Salaries, summary)
```

```
##      rank    sex salary.Min. salary.1st Qu. salary.Median salary.Mean
## 1 AssocProf Female      62880          73680          90560      88510
## 2 AsstProf  Female      63100          73000          77000      78050
## 3 Prof     Female      90450         109700         120300     122000
## 4 AssocProf  Male      70000          83210          95630      94870
## 5 AsstProf   Male      63900          74380          80180      81310
## 6 Prof      Male      57800         105800         124000     127100
## salary.3rd Qu. salary.Max.
## 1         103900      109600
## 2          79360       97030
## 3         135000      161100
## 4         104600      126400
## 5          88800       95080
## 6         145500      231500
```

However, we are not limited to just rank and sex. We can also look at how age and rank affect salary

```
aggregate(salary ~ yrs.service + rank, Salaries, summary)
```

##	yrs.service	rank	salary.Min.	salary.1st Qu.	salary.Median
## 1	1	AssocProf	104800	106600	108400
## 2	5	AssocProf	83900	84650	85400
## 3	6	AssocProf	88650	96200	101200
## 4	7	AssocProf	70000	95440	105100
## 5	8	AssocProf	74830	81650	86610
## 6	9	AssocProf	71060	86160	97890
## 7	10	AssocProf	81500	83850	99250
## 8	11	AssocProf	103600	103700	103800
## 9	12	AssocProf	95610	95610	95610
## 10	16	AssocProf	82100	82100	82100
## 11	17	AssocProf	81280	81280	81280
## 12	18	AssocProf	113300	113300	113300
## 13	19	AssocProf	86250	86250	86250
## 14	22	AssocProf	62880	73500	84120
## 15	23	AssocProf	74000	78850	83710
## 16	24	AssocProf	73300	73300	73300
## 17	28	AssocProf	106300	106300	106300
## 18	33	AssocProf	88600	88600	88600
## 19	39	AssocProf	70700	70700	70700
## 20	49	AssocProf	81800	81800	81800
## 21	53	AssocProf	90000	90000	90000
## 22	0	AsstProf	72500	77000	81000
## 23	1	AsstProf	63900	72500	72750
## 24	2	AsstProf	73000	79220	80220
## 25	3	AsstProf	68400	73720	77790
## 26	4	AsstProf	74000	79910	86370
## 27	5	AsstProf	74000	82610	91230
## 28	6	AsstProf	63100	63100	63100
## 29	0	Prof	105000	105000	105000
## 30	2	Prof	96540	111400	126300
## 31	3	Prof	117200	125500	133800
## 32	4	Prof	105300	112000	118800
## 33	5	Prof	141100	147200	153300
## 34	6	Prof	93000	94600	96200
## 35	7	Prof	92050	107300	116400
## 36	8	Prof	102000	104900	106100
## 37	9	Prof	106600	111200	116500
## 38	10	Prof	104400	106000	107700
## 39	11	Prof	88180	106400	119500
## 40	12	Prof	101000	121400	130100
## 41	13	Prof	170500	170500	170500
## 42	14	Prof	102200	107600	112600
## 43	15	Prof	95330	109600	124700
## 44	16	Prof	112400	130800	135600
## 45	17	Prof	81700	111500	124300
## 46	18	Prof	101100	120500	127600
## 47	19	Prof	94350	114200	146900
## 48	20	Prof	101000	120000	126600
## 49	21	Prof	117700	120500	135100

## 50	22	Prof	96610	101600	124100
## 51	23	Prof	84270	101900	119100
## 52	24	Prof	93160	93160	93160
## 53	25	Prof	101700	113400	130800
## 54	26	Prof	89560	105200	121200
## 55	27	Prof	91000	108100	139200
## 56	28	Prof	98190	117600	124600
## 57	29	Prof	148500	148500	148500
## 58	30	Prof	92550	103100	122900
## 59	31	Prof	99420	110200	125600
## 60	32	Prof	124300	124300	124300
## 61	33	Prof	128200	145200	162200
## 62	34	Prof	92390	95160	97920
## 63	35	Prof	87800	99000	100400
## 64	36	Prof	88600	92850	107300
## 65	37	Prof	102600	103900	128000
## 66	38	Prof	93520	129100	143600
## 67	39	Prof	109000	110200	111400
## 68	40	Prof	77200	88710	101000
## 69	41	Prof	141500	141500	141500
## 70	43	Prof	72300	126000	149900
## 71	44	Prof	89650	97320	105000
## 72	45	Prof	67560	97550	127200
## 73	46	Prof	100600	100600	100600
## 74	48	Prof	107200	107200	107200
## 75	49	Prof	78160	105400	132600
## 76	51	Prof	57800	57800	57800
## 77	57	Prof	76840	76840	76840
## 78	60	Prof	192300	192300	192300
##	salary.Mean	salary.3rd Qu.	salary.Max.		
## 1	110600	113600	118700		
## 2	85400	86150	86900		
## 3	99850	104800	107200		
## 4	97650	107000	113600		
## 5	90340	100100	119800		
## 6	91950	100700	100900		
## 7	94470	103800	104000		
## 8	111300	115100	126400		
## 9	95610	95610	95610		
## 10	82100	82100	82100		
## 11	81280	81280	81280		
## 12	113300	113300	113300		
## 13	86250	86250	86250		
## 14	84120	94730	105400		
## 15	83710	88560	93420		
## 16	73300	73300	73300		
## 17	106300	106300	106300		
## 18	88600	88600	88600		
## 19	70700	70700	70700		
## 20	81800	81800	81800		
## 21	90000	90000	90000		
## 22	81630	87250	92000		
## 23	74740	76990	86100		
## 24	82350	88510	91300		

## 25	79710	86020	95080
## 26	85510	92000	92700
## 27	87420	94130	97030
## 28	63100	63100	63100
## 29	105000	105000	105000
## 30	123100	136400	146500
## 31	133800	142100	150500
## 32	118800	125500	132300
## 33	153100	159200	165000
## 34	112000	121500	146800
## 35	129500	129700	204000
## 36	108100	109300	118200
## 37	128700	120800	183800
## 38	114300	113600	145200
## 39	120400	143900	148800
## 40	126500	135200	145000
## 41	170500	170500	170500
## 42	118600	128800	147300
## 43	128200	137300	166800
## 44	141000	152200	173200
## 45	126100	152500	160400
## 46	137400	149400	194800
## 47	137100	151700	193000
## 48	128600	137000	163200
## 49	138800	153100	170000
## 50	122100	138600	150000
## 51	121300	133300	175000
## 52	93160	93160	93160
## 53	133800	147200	172300
## 54	122600	140700	155500
## 55	129000	146500	163200
## 56	130400	145900	168500
## 57	148500	148500	148500
## 58	118400	134000	138800
## 59	126500	133800	162200
## 60	124300	124300	124300
## 61	160000	175800	189400
## 62	97920	100700	103400
## 63	109000	107300	150400
## 64	108500	119000	137000
## 65	127800	151900	152700
## 66	147300	155100	231500
## 67	111800	113200	115000
## 68	106000	119700	143200
## 69	141500	141500	141500
## 70	144400	168300	205500
## 71	112900	124500	144000
## 72	117400	147100	147800
## 73	100600	100600	100600
## 74	107200	107200	107200
## 75	132600	159800	187000
## 76	57800	57800	57800
## 77	76840	76840	76840
## 78	192300	192300	192300

Finally, it is equally as simple to look at how salary may depend on sex and discipline

```
aggregate(salary ~ sex + discipline, Salaries, summary)
```

```
##      sex discipline salary.Min. salary.1st Qu. salary.Median salary.Mean
## 1 Female          A      62880      73350      78000      89060
## 2  Male          A      57800      87350     105300     110700
## 3 Female          B      71060      97030     105400     111200
## 4  Male          B      67560      94730     113600     118800
## salary.3rd Qu. salary.Max.
## 1      108500      137000
## 2      128000      205500
## 3      127500      161100
## 4      141300      231500
```