

R Notebook

```
# Name: Arjun Bhan
library("tidyverse")
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.0.6      v dplyr   1.0.4
## v tidyr   1.1.2      v stringr 1.4.0
## v readr   1.4.0      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
AstroMil <- read.csv(file.choose(),stringsAsFactors = FALSE)
AstroCiv <-read.csv(file.choose(),stringsAsFactors = FALSE)
```

AstroCiv

id	number	nationwide_number	name	
<int>	<int>	<int>	<chr>	
21	12	8	Feoktistov, Konstantin	
21	12	8	Feoktistov, Konstantin	
22	13	9	Yegorov, Boris	
22	13	9	Yegorov, Boris	
50	24	13	Armstrong, Neil A.	
50	24	13	Armstrong, Neil A.	
51	24	13	Armstrong, Neil A.	
51	24	13	Armstrong, Neil A.	
58	27	16	Collins, Michael	
58	27	16	Collins, Michael	

1-10 of 1,016 rows | 1-4 of 23 columns

Previous 1 2 3 4 5 6 ... 102 Next

AstroMil

id <int>	number <int>	nationwide_number <int>	name <chr>
1	1	1	Gagarin, Yuri
2	2	2	Titov, Gherman
3	3	1	Glenn, John H., Jr.
4	3	1	Glenn, John H., Jr.
5	4	2	Carpenter, M. Scott
6	5	2	Nikolayev, Andriyan
7	5	2	Nikolayev, Andriyan
8	6	4	Popovich, Pavel
9	6	4	Popovich, Pavel
10	7	3	Schirra, Walter M., Jr.

1-10 of 769 rows | 1-4 of 23 columns

Previous 1 2 3 4 5 6 ... 77 Next

This data is not tidy as many of the rows share the same information. By using the spread comm and we are able to make sure that the rows contain separate information. I plan on using the spr ead command on the eva column's values. With these values separated it will help the civilian da ta base have less shared information between the individual rows and therefore, be more tidy.

```
AstroCivTidy <- AstroCiv %>% spread(key="eva",value="hours")
AstroCivTidy["military_civilian"]= "civilian"
AstroCivTidy
```

id <int>	number <int>	nationwide_number <int>	name <chr>
21	12	8	Feoktistov, Konstantin
22	13	9	Yegorov, Boris
50	24	13	Armstrong, Neil A.
51	24	13	Armstrong, Neil A.
58	27	16	Collins, Michael
59	27	16	Collins, Michael
65	31	20	Cunningham, Ronnie Walter
66	100	50	Savinykh, Viktor
67	100	50	Savinykh, Viktor
68	101	1	Jugderdemidiin Gurragchaa

1-10 of 508 rows | 1-4 of 24 columns

Previous 1 2 3 4 5 6 ... 51 Next

```
AstroMil["military_civilian"] = "military"
AstroMil
```

id <int>	number <int>	nationwide_number <int>	name <chr>	
1	1	1	Gagarin, Yuri	
2	2	2	Titov, Gherman	
3	3	1	Glenn, John H., Jr.	
4	3	1	Glenn, John H., Jr.	
5	4	2	Carpenter, M. Scott	
6	5	2	Nikolayev, Andriyan	
7	5	2	Nikolayev, Andriyan	
8	6	4	Popovich, Pavel	
9	6	4	Popovich, Pavel	
10	7	3	Schirra, Walter M., Jr.	

1-10 of 769 rows | 1-4 of 24 columns

Previous 1 2 3 4 5 6 ... 77 Next

```
FullAstroTab<- full_join(AstroMil, AstroCivTidy)
```

```
## Joining, by = c("id", "number", "nationwide_number", "name", "original_name", "sex", "year_of_
_birth", "nationality", "selection", "year_of_selection", "mission_number", "total_number_of_mis
sions", "occupation", "year_of_mission", "mission_title", "ascend_shuttle", "in_orbit", "descend
_shuttle", "hours_mission", "total_hrs_sum", "field21", "eva_hrs_mission", "total_eva_hrs", "mil
itary_civilian")
```

```
FullAstroTab
```

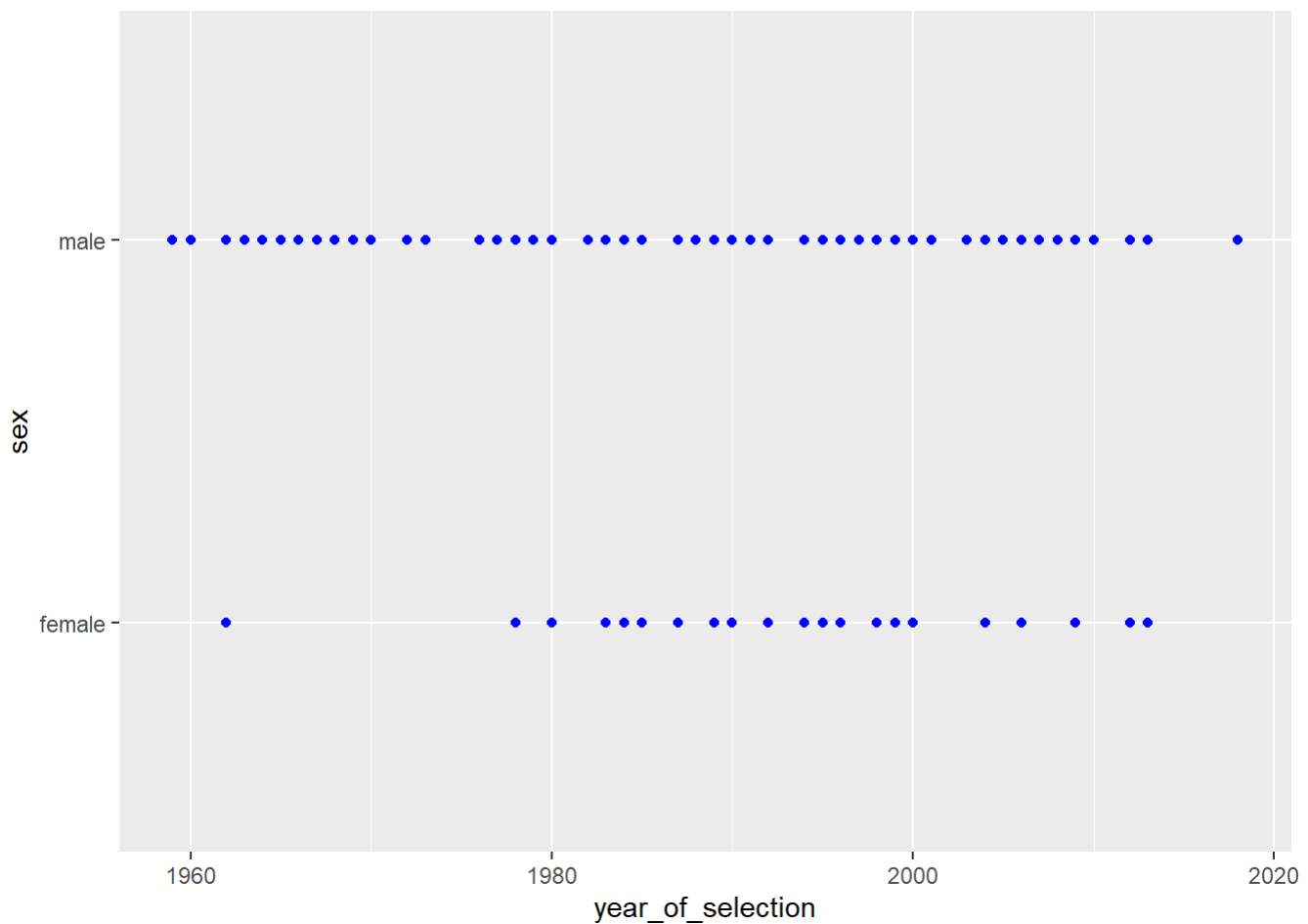
id <int>	number <int>	nationwide_number <int>	name <chr>	
1	1	1	Gagarin, Yuri	
2	2	2	Titov, Gherman	
3	3	1	Glenn, John H., Jr.	
4	3	1	Glenn, John H., Jr.	
5	4	2	Carpenter, M. Scott	
6	5	2	Nikolayev, Andriyan	
7	5	2	Nikolayev, Andriyan	

id <int>	number <int>	nationwide_number <int>	name <chr>
8	6	4	Popovich, Pavel
9	6	4	Popovich, Pavel
10	7	3	Schirra, Walter M., Jr.

1-10 of 1,277 rows | 1-4 of 24 columns

Previous 1 2 3 4 5 6 ... 128 Next

```
ggplot(FullAstroTab, aes(x=year_of_selection, y=sex))+geom_point(color="blue")
```



#This graph shows two key facts about the data. First, there have been far more men than woman chosen to be astronauts. This also shows that it was uncommon for the first 15 years for woman to become astronauts. After many years, woman astronauts have become more common.