HW 4 - IE 6600 - Sec 03 - Team 5

Rahul Dixit, Kirti Aggarwal, Asif Khan

2/24/2021

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

## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.0.5      v dplyr   1.0.3
## 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()

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

## corrplot 0.84 loaded

##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine

##
## Attaching package: 'scales'

## The following object is masked from 'package:purrr':
##
##     discard

## The following object is masked from 'package:readr':
##
##     col_factor

##
## Attaching package: 'magrittr'
```

```
## The following object is masked from 'package:purrr':
##
##     set_names

## The following object is masked from 'package:tidyr':
##
##     extract


##
## -- Column specification --------------------------------------------------
## cols(
##   year = col_double(),
##   month = col_double(),
##   carrier = col_character(),
##   carrier_name = col_character(),
##   airport = col_character(),
##   airport_name = col_character(),
##   arr_flights = col_double(),
##   arr_del15 = col_double(),
##   arr_cancelled = col_double(),
##   arr_diverted = col_double(),
##   arr_delay = col_double(),
##   carrier_delay = col_double(),
##   weather_delay = col_double(),
##   nas_delay = col_double(),
##   security_delay = col_double(),
##   late_aircraft_delay = col_double()
## )


##
## -- Column specification --------------------------------------------------
## cols(
##   `ID Gender` = col_double(),
##   Gender = col_character(),
##   `ID Year` = col_double(),
##   Year = col_double(),
##   `ID Workforce Status` = col_logical(),
##   `Workforce Status` = col_logical(),
##   `Record Count` = col_double(),
##   `Average Wage` = col_double(),
##   `Average Wage Appx MOE` = col_double(),
##   `Total Population` = col_double(),
##   `Total Population MOE Appx` = col_double(),
##   `PUMS Occupation` = col_character(),
##   `ID PUMS Occupation` = col_double(),
##   `Slug PUMS Occupation` = col_character(),
##   `PUMS Industry` = col_character(),
##   `ID PUMS Industry` = col_double(),
##   `Slug PUMS Industry` = col_character()
## )


##
## -- Column specification --------------------------------------------------
```

```
## cols(
##   .default = col_character(),
##   'ID Year' = col_double(),
##   Year = col_double(),
##   'ID Industry Sub-Sector' = col_double(),
##   'ID Workforce Status' = col_logical(),
##   'Workforce Status' = col_logical(),
##   'Total Population' = col_double(),
##   'Total Population MOE Appx' = col_double(),
##   'Average Wage' = col_double(),
##   'Average Wage Appx MOE' = col_double(),
##   'Record Count' = col_double()
## )
## i Use 'spec()' for the full column specifications.
```

Task 1

Problem Statement

Create a density plot for all different airlines delays by calling a custom function that has transformed x-axis scale for better visualisation.

Result
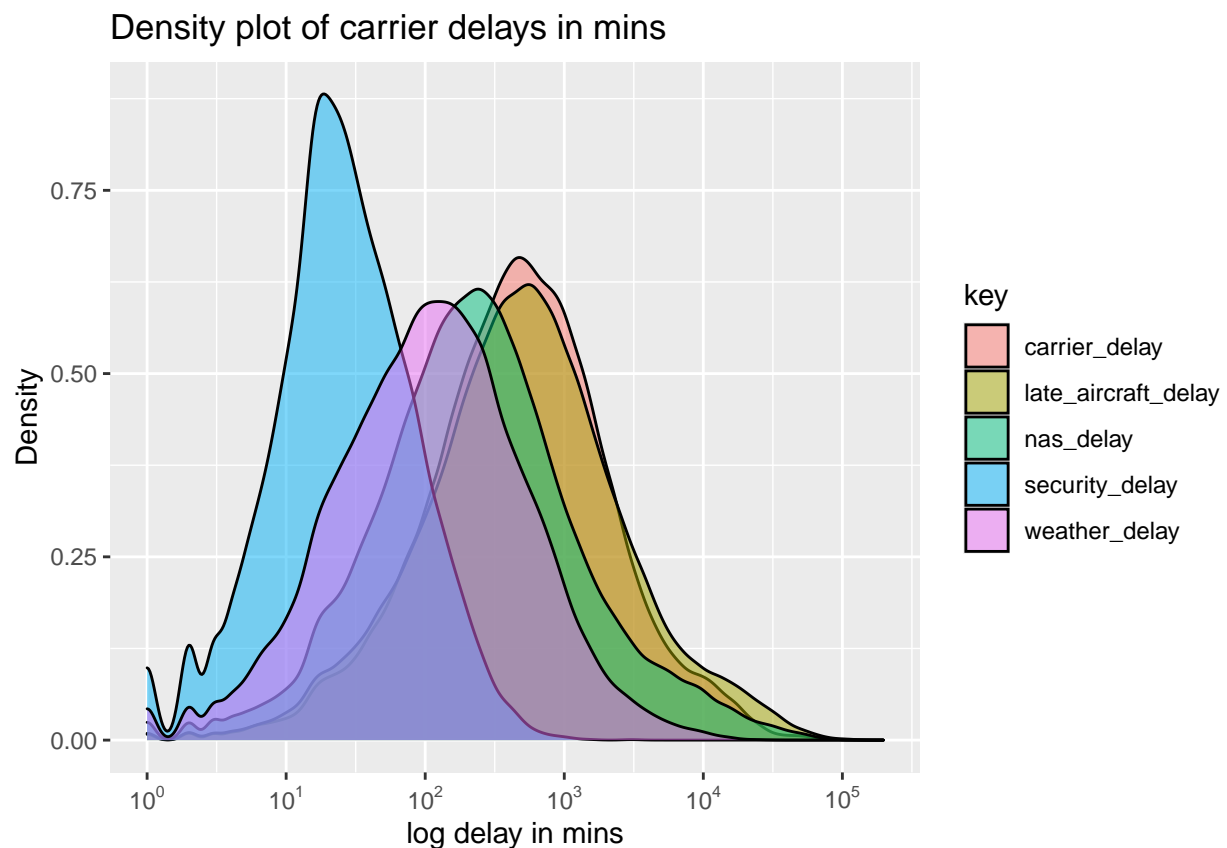
Conclusion

From the plot we can see that security delay has the highest density and the carrier delay has a higher spread compared to the rest of the delays

Task 2

Problem Statement

Create a correlation plot for selected columns (arr_flights, arr_del15, arr_cancelled, arr_diverted, arr_delay, carrier_delay, weather_delay, nas_delay, security_delay and late_aircraft_delay) having labels properly aligned.

Result
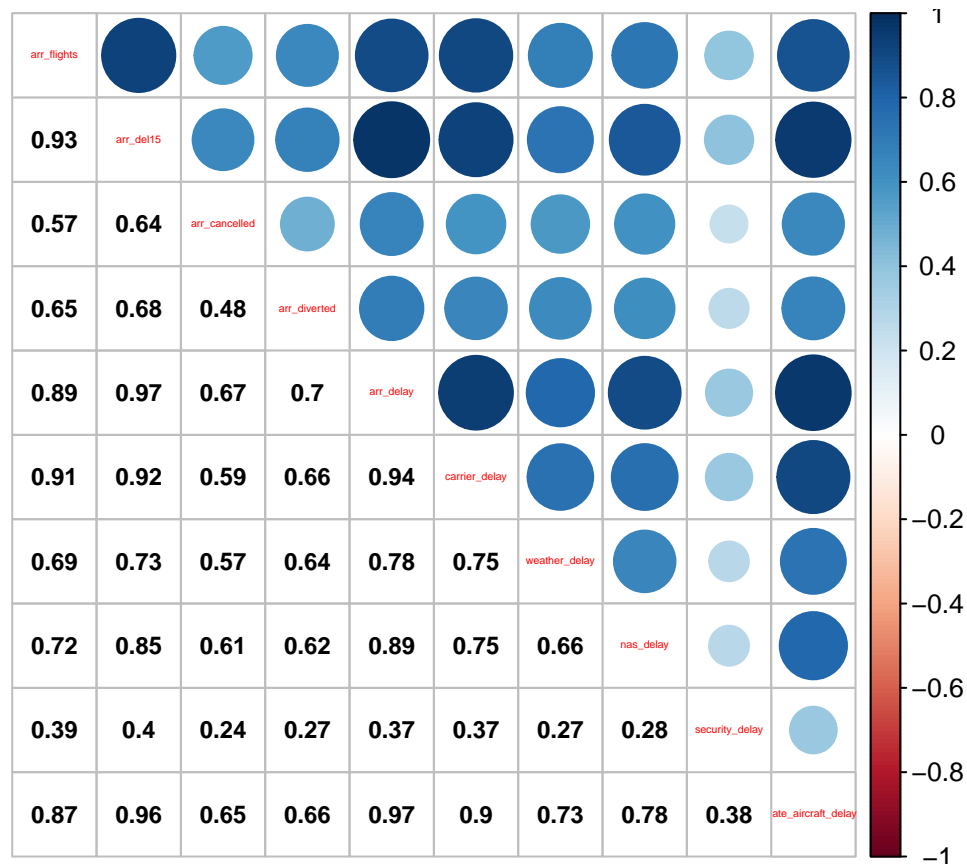
Conclusion

From the graph we can conclude that aircraft delay and total delay is the most correlated feature with the correlation value of 0.97.The least correlation is between canceled delay and security delay with a value of 0.24.

Task 3

Problem statement

Create a heatmap to find difference between male and female employees w.r.t to year and occupation

Result
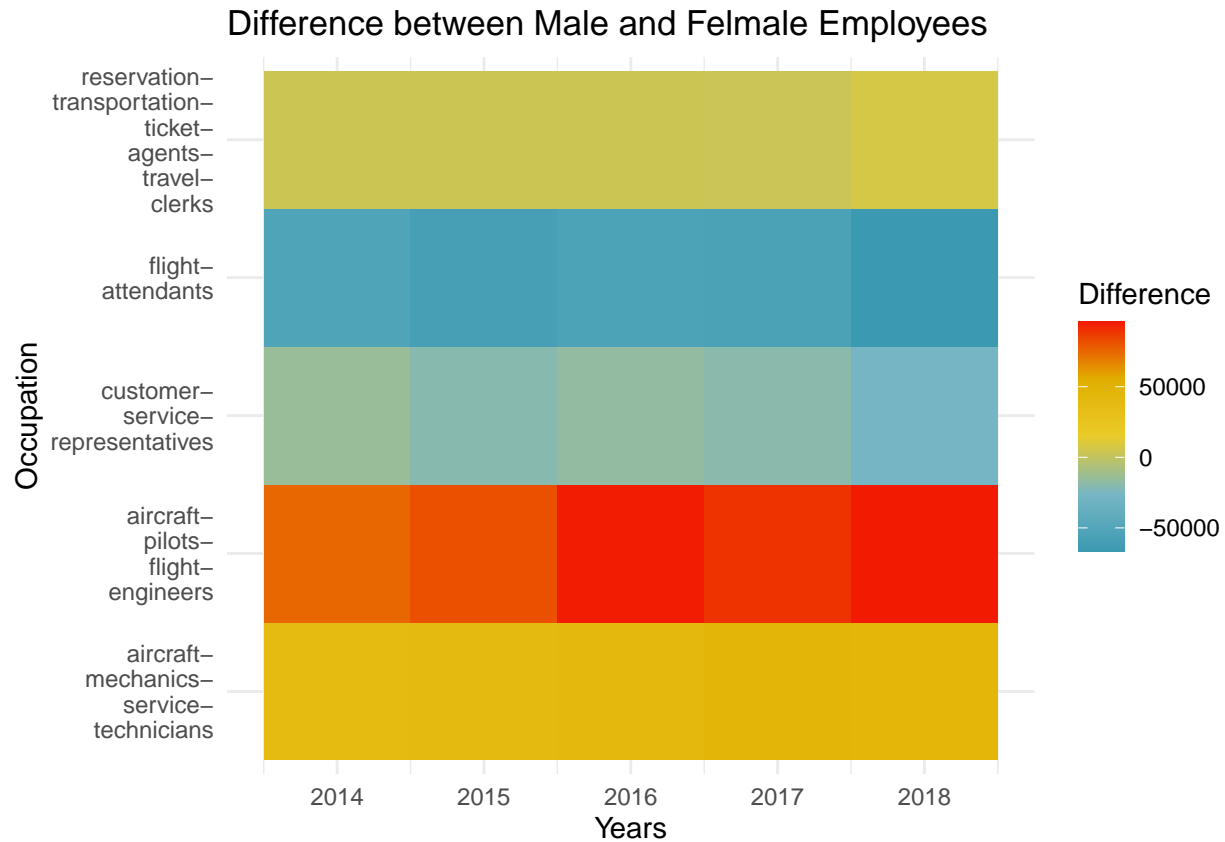
Conclusion

As clearly seen from the graph there for the occupation of

1. Flight attendants, there are more female employees than male employees
2. Aircraft mechanics and Reservation Transportation, there are almost same number of male and female employees
3. Aircraft pilots, there are more male employees than female employees

Task 4

Problem statement

Create an Alluvial Chart for Occupation and Year w.r.t Population, showing inter-relation of Gender.
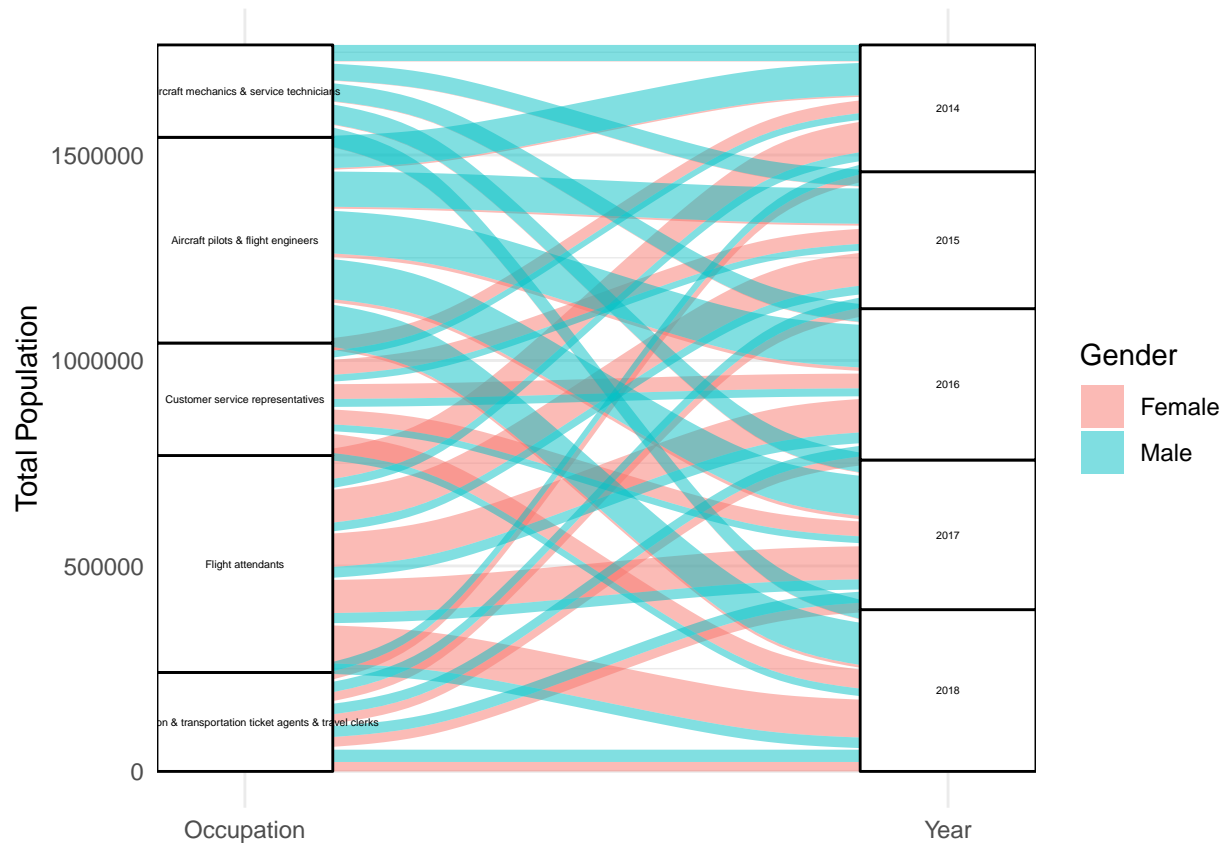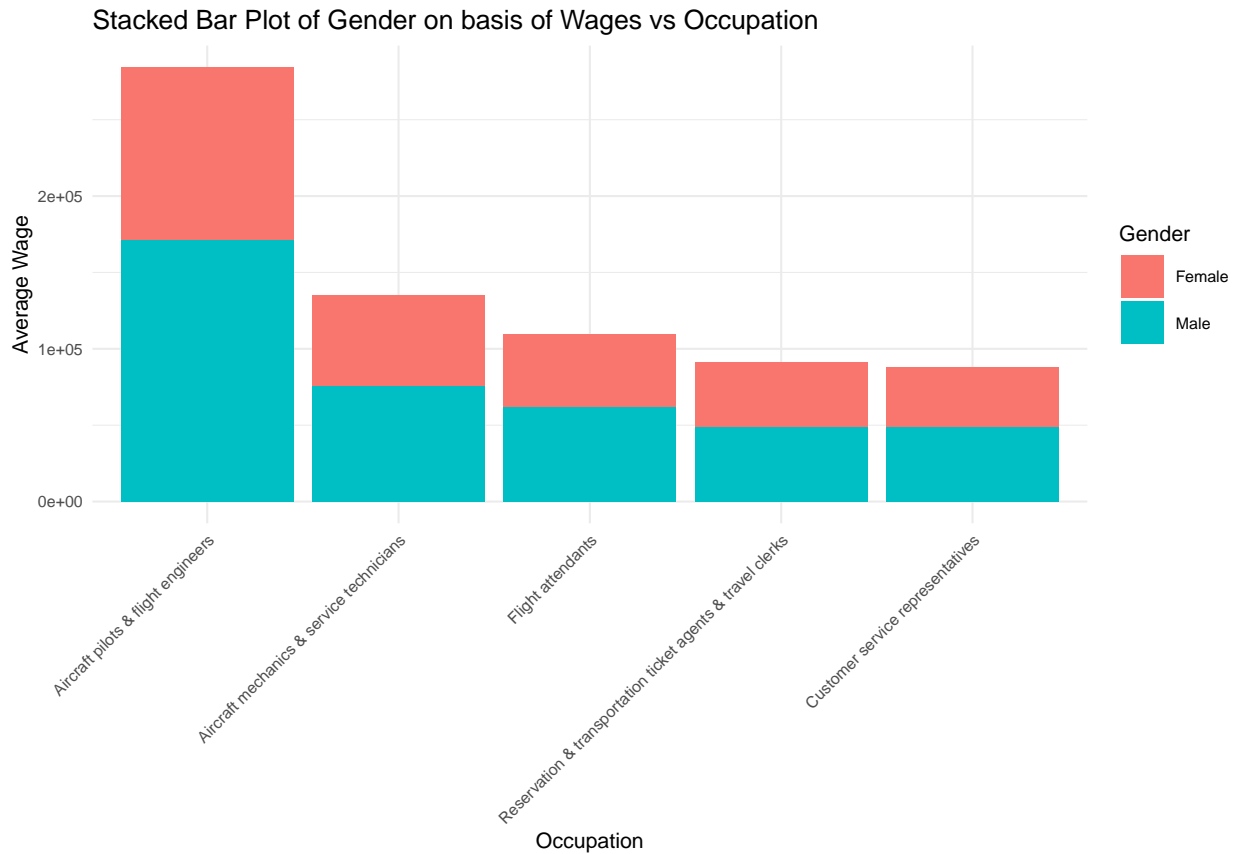
Conclusion

1. There are more male employees from the year 2014 to 2018 having an occupation in Aircraft Mechanics, Pilots.
2. There are more female employees from the year 2014 to 2018 having an occupation in Customer service, Flight Attendants and reservation transportation.

Task 5

Problem statement

Create a stacked bar plot of Occupation vs Wage (in year 2018) on the basis of Gender as different color

Result

Stacked Bar Plot of Gender on basis of Wages vs Occupation

Conclusion

1. Average wage for different occupation is more for Males as compared to females for 3 categories (Aircraft Pilots and Flight engineers, Aircraft Mechanics and service technicians Flight Attendants) out of 5. Whereas 2 categories have comparable wage for both males and females.
2. Aircraft Pilots and Flight engineers has the highest average wage.
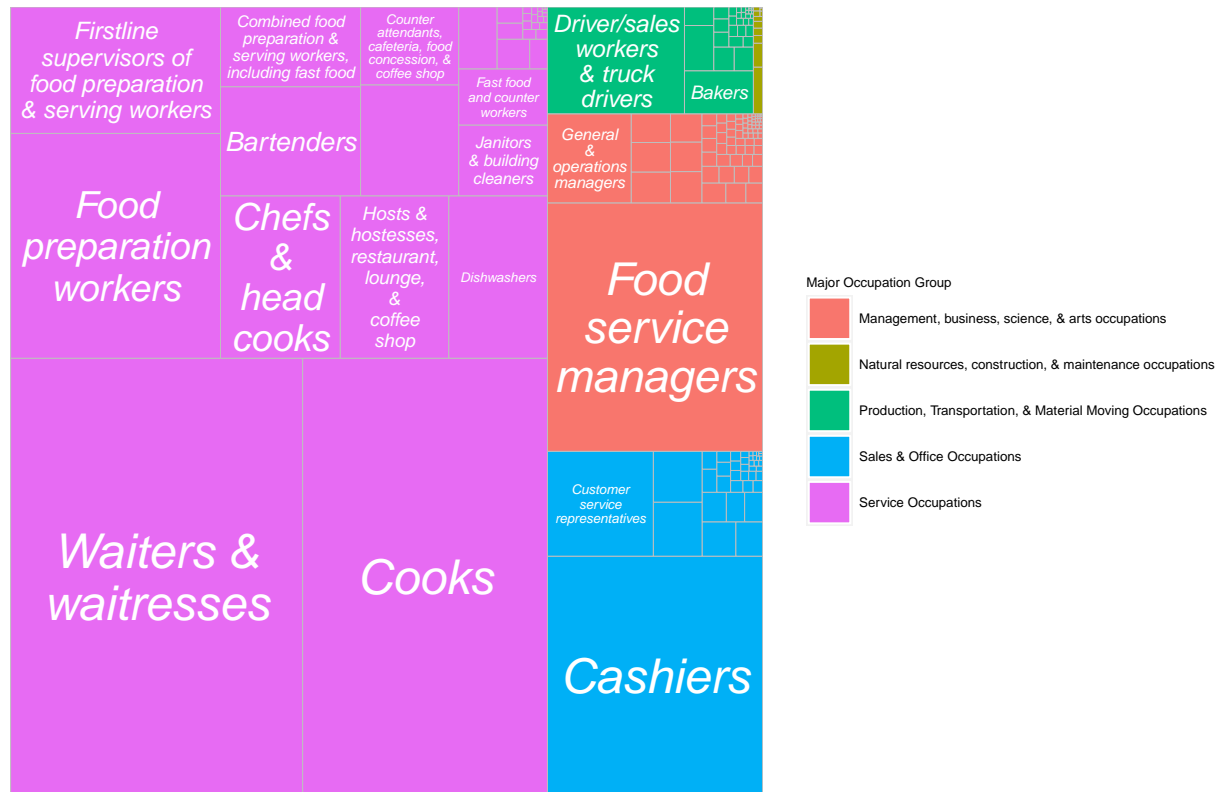3. Customer services representatives has the lowest average wage.

Task 6

Problem statement

Create a tree map in which area of each rectangle (Major Occupation Group) is proportional to the number of people working in that Detailed Occupation.

Result

Workforce Distribution by Detailed Workforce 2018

Conclusion

1. Service Occupation is the subgroup where maximum number of people work.
2. Natural Resources, construction and maintenance occupation is the subgroup where least number of people work.