Assignment 7 - CT5102

Relational Data with dplyr

The aim of this exercise is to analyse time series Dublin Airport passenger data, which was extracted from the Central Statistics Office. The data is in wide data format, and contains in two workskeets. *DublinInward* are flights coming into Dublin, while *DublinOutward* are outbound flights. The first 10 rows and columns for the data is as follows:

```
d_in[1:10,1:10]
```

```
## # A tibble: 10 x 10
##    Date  `Aberdeen (ABZ)~ `Lanzarote (ACE~ `Izmir (ADB),Tu~
##    <chr>            <dbl>            <dbl>            <dbl>
##  1 2006~             3041            10075              130
##  2 2006~             3316             8576                0
##  3 2006~             4562             9295                0
##  4 2006~             4172            10398              980
##  5 2006~             3836             8688             2385
##  6 2006~             4039            11420             2741
##  7 2006~             4990            15030             4848
##  8 2006~             4520            14213             4121
##  9 2006~             3929            12368             3805
## 10 2006~             4182            10579             2243
## # ... with 6 more variables: `Agadir (AGA),Morocco` <dbl>, `Malaga
## #   (AGP),Spain` <dbl>, `Alghero (AHO),Italy` <dbl>, `Alicante
## #   (ALC),Spain` <dbl>, `Amsterdam (AMS),Netherlands` <dbl>, `Stockholm -
## #   Arlanda (ARN),Sweden` <dbl>
```

```
d_out[1:10,1:10]
```

```
## # A tibble: 10 x 10
##    Date  `Aberdeen (ABZ)~ `Lanzarote (ACE~ `Izmir (ADB),Tu~
##    <chr>            <dbl>            <dbl>            <dbl>
##  1 2006~             3504             9301                0
##  2 2006~             3383             8797                0
##  3 2006~             4115             9102                0
##  4 2006~             4308            10553             1543
##  5 2006~             3517             9565             2874
##  6 2006~             3940            12585             3293
##  7 2006~             4677            15295             5170
##  8 2006~             4576            12582             3734
##  9 2006~             4099            12143             3410
## 10 2006~             4348             9322             1644
## # ... with 6 more variables: `Agadir (AGA),Morocco` <dbl>, `Malaga
## #   (AGP),Spain` <dbl>, `Alghero (AHO),Italy` <dbl>, `Alicante
## #   (ALC),Spain` <dbl>, `Amsterdam (AMS),Netherlands` <dbl>, `Stockholm -
## #   Arlanda (ARN),Sweden` <dbl>
```

Use the appropriate functions to transform the input data sets to their tidy data equivalent.

```
d_in_tidy
```

```
## # A tibble: 6,923 x 7
##    Year  Month Airport             Passengers Direction MonthName Date
##    <chr> <chr> <chr>                    <int> <chr>     <fct>     <date>
##  1 2006  01    Aberdeen (ABZ),Gr~        3041 Inbound   Jan       2006-01-01
##  2 2006  02    Aberdeen (ABZ),Gr~        3316 Inbound   Feb       2006-02-01
##  3 2006  03    Aberdeen (ABZ),Gr~        4562 Inbound   Mar       2006-03-01
##  4 2006  04    Aberdeen (ABZ),Gr~        4172 Inbound   Apr       2006-04-01
##  5 2006  05    Aberdeen (ABZ),Gr~        3836 Inbound   May       2006-05-01
##  6 2006  06    Aberdeen (ABZ),Gr~        4039 Inbound   Jun       2006-06-01
##  7 2006  07    Aberdeen (ABZ),Gr~        4990 Inbound   Jul       2006-07-01
##  8 2006  08    Aberdeen (ABZ),Gr~        4520 Inbound   Aug       2006-08-01
##  9 2006  09    Aberdeen (ABZ),Gr~        3929 Inbound   Sep       2006-09-01
## 10 2006  10    Aberdeen (ABZ),Gr~        4182 Inbound   Oct       2006-10-01
## # ... with 6,913 more rows
```

```
d_out_tidy
```

```
## # A tibble: 6,923 x 7
##    Year  Month Airport             Passengers Direction MonthName Date
##    <chr> <chr> <chr>                    <int> <chr>     <fct>     <date>
##  1 2006  01    Aberdeen (ABZ),Gr~        3504 Outbound  Jan       2006-01-01
##  2 2006  02    Aberdeen (ABZ),Gr~        3383 Outbound  Feb       2006-02-01
##  3 2006  03    Aberdeen (ABZ),Gr~        4115 Outbound  Mar       2006-03-01
##  4 2006  04    Aberdeen (ABZ),Gr~        4308 Outbound  Apr       2006-04-01
##  5 2006  05    Aberdeen (ABZ),Gr~        3517 Outbound  May       2006-05-01
##  6 2006  06    Aberdeen (ABZ),Gr~        3940 Outbound  Jun       2006-06-01
##  7 2006  07    Aberdeen (ABZ),Gr~        4677 Outbound  Jul       2006-07-01
##  8 2006  08    Aberdeen (ABZ),Gr~        4576 Outbound  Aug       2006-08-01
##  9 2006  09    Aberdeen (ABZ),Gr~        4099 Outbound  Sep       2006-09-01
## 10 2006  10    Aberdeen (ABZ),Gr~        4348 Outbound  Oct       2006-10-01
## # ... with 6,913 more rows
```

Create a full tidy data set encompassing both the inbound and outbound data.

```
full_data
```

```
## # A tibble: 13,846 x 7
##    Year  Month Airport             Passengers Direction MonthName Date
##    <chr> <chr> <chr>                    <int> <chr>     <fct>     <date>
##  1 2006  01    Aberdeen (ABZ),Gr~        3041 Inbound   Jan       2006-01-01
##  2 2006  02    Aberdeen (ABZ),Gr~        3316 Inbound   Feb       2006-02-01
##  3 2006  03    Aberdeen (ABZ),Gr~        4562 Inbound   Mar       2006-03-01
##  4 2006  04    Aberdeen (ABZ),Gr~        4172 Inbound   Apr       2006-04-01
##  5 2006  05    Aberdeen (ABZ),Gr~        3836 Inbound   May       2006-05-01
##  6 2006  06    Aberdeen (ABZ),Gr~        4039 Inbound   Jun       2006-06-01
##  7 2006  07    Aberdeen (ABZ),Gr~        4990 Inbound   Jul       2006-07-01
##  8 2006  08    Aberdeen (ABZ),Gr~        4520 Inbound   Aug       2006-08-01
##  9 2006  09    Aberdeen (ABZ),Gr~        3929 Inbound   Sep       2006-09-01
## 10 2006  10    Aberdeen (ABZ),Gr~        4182 Inbound   Oct       2006-10-01
## # ... with 13,836 more rows
```

In terms of passenger numbers (to and from Dublin Airports), list the top ten.

```
top_ten
```

```
##  [1] "Birmingham - Uk (BHX),Great Britain"
##  [2] "Paris - Charles De Gaulle (CDG),France"
##  [3] "Amsterdam (AMS),Netherlands"
##  [4] "Malaga (AGP),Spain"
##  [5] "Barcelona (BCN),Spain"
##  [6] "Bristol (BRS),Great Britain"
##  [7] "Brussels (BRU),Belgium"
##  [8] "Lanzarote (ACE),Spain"
##  [9] "Boston (BOS),USA"
## [10] "Copenhagen (CPH),Denmark"
```

Summarise the Yearly values, by direction

```
summ_year
```

```
## # A tibble: 28 x 3
## # Groups:   Year [14]
##    Year  Direction   Total
##    <chr> <chr>       <int>
##  1 2006  Inbound   2611144
##  2 2006  Outbound  2608234
##  3 2007  Inbound   2828018
##  4 2007  Outbound  2808869
##  5 2008  Inbound   2896901
##  6 2008  Outbound  2880994
##  7 2009  Inbound   2666104
##  8 2009  Outbound  2658559
##  9 2010  Inbound   2551237
## 10 2010  Outbound  2553728
## # ... with 18 more rows
```

Summarise the average monthly values, by direction

```
avr_month
```

```
## # A tibble: 322 x 5
## # Groups:   Year, Month [161]
##    Year  Month Direction Total Date
##    <chr> <chr> <chr>     <dbl> <date>
##  1 2006  01    Inbound   3940. 2006-01-01
##  2 2006  01    Outbound  3629. 2006-01-01
##  3 2006  02    Inbound   4010. 2006-02-01
##  4 2006  02    Outbound  3941. 2006-02-01
##  5 2006  03    Inbound   4451. 2006-03-01
##  6 2006  03    Outbound  4487. 2006-03-01
##  7 2006  04    Inbound   5121. 2006-04-01
##  8 2006  04    Outbound  5155. 2006-04-01
##  9 2006  05    Inbound   5402. 2006-05-01
## 10 2006  05    Outbound  5617. 2006-05-01
## # ... with 312 more rows
```

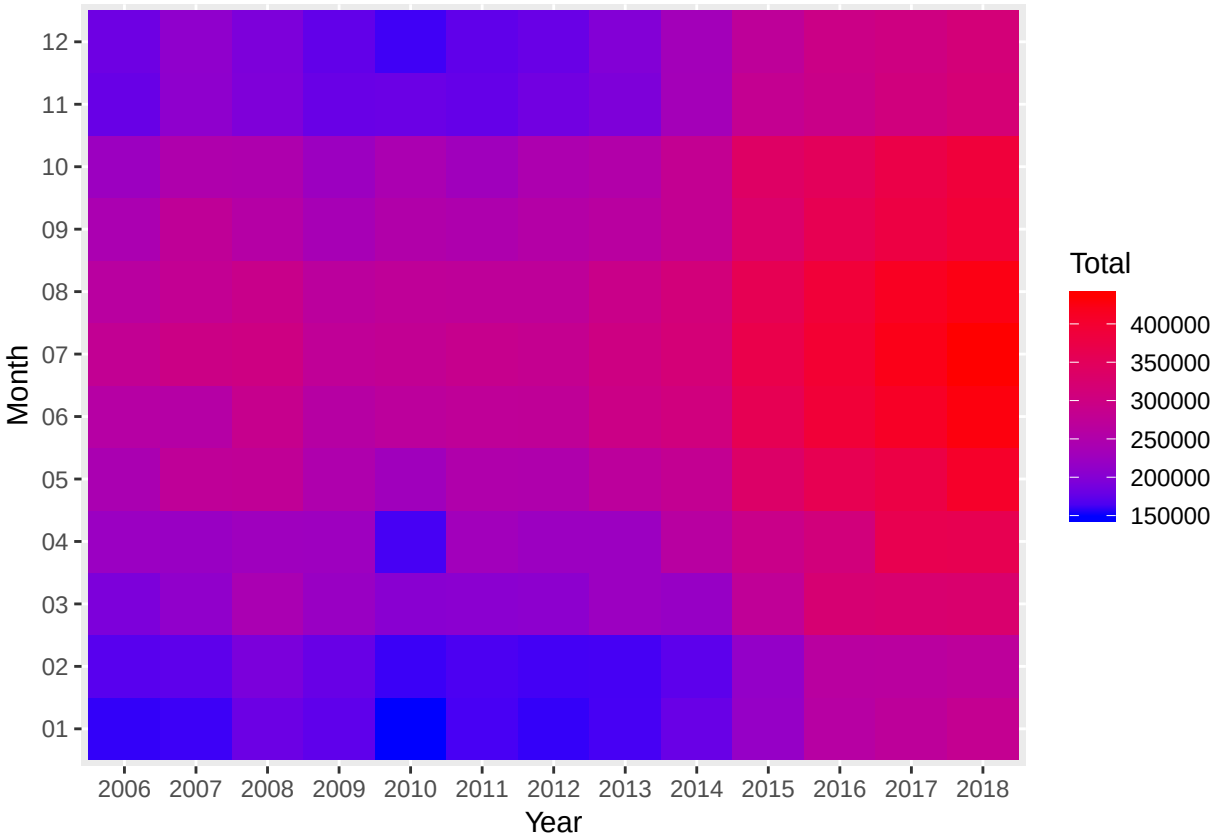
Find the total outbound passengers per month (excluding 2019):

```
Total_Pass
```

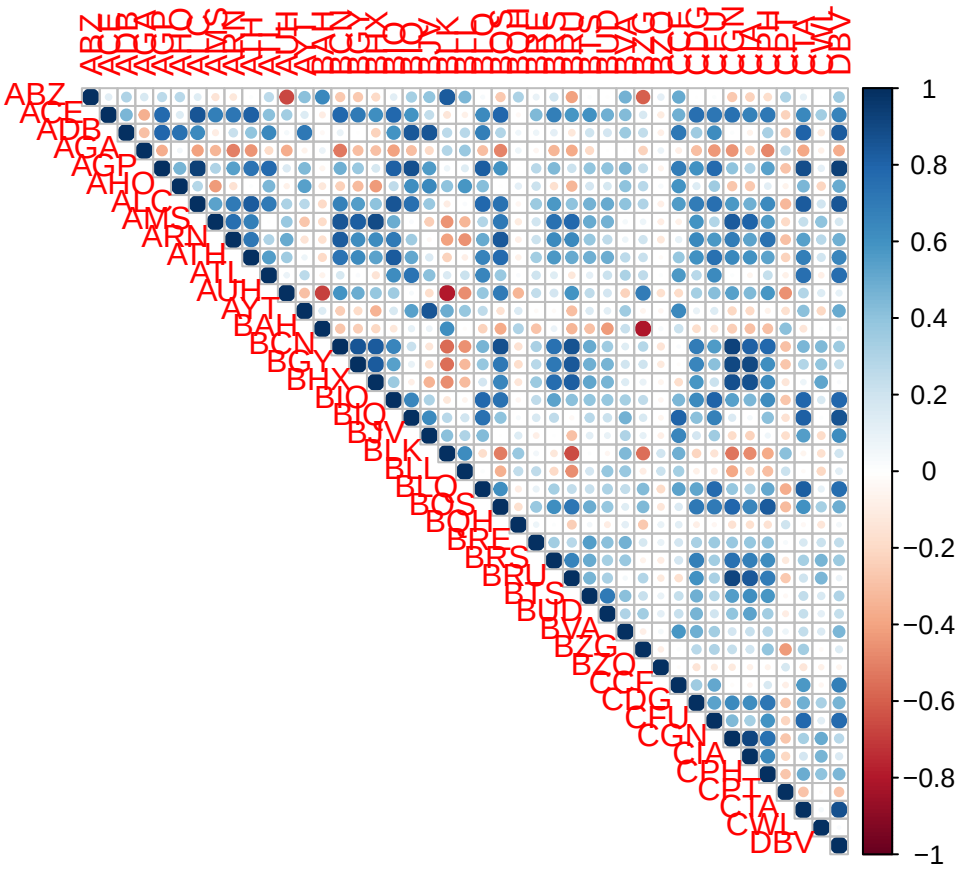
```
## # A tibble: 156 x 3
```

```
## # Groups:   Year [13]
##    Year  Month  Total
##    <chr> <chr>  <int>
##  1 2006  01    156030
##  2 2006  02    169452
##  3 2006  03    192947
##  4 2006  04    221668
##  5 2006  05    241517
##  6 2006  06    257914
##  7 2006  07    279058
##  8 2006  08    263339
##  9 2006  09    242117
## 10 2006  10    223772
## # ... with 146 more rows
```

Display a heat map of passenger numbers by month. Use the functions **geom_tile()** and **scale_fill_gradient()**



Based on the outward passenger numbers for each destination, create the following correlation plot matrix (using the package **corrplot**)


ABZ
ACE
ADB
AGA
AGP
AHO
ALC
AMS
APN
ATH
ATL
AUH
AYT
BAH
BCN
BGY
BHX
BIO
BIQ
BJV
BLK
BLL
BLQ
BOS
BQH
BRE
BRS
BRU
BTS
BUD
BVA
BZG
BZQ
CCF
CDG
CFU
CGN
CIA
CPH
CPT
CTA
CWL
DBV
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1