

Data Science for Operational Researchers Using R ONLINE



Tutor: Dr James Duggan

Session 1 - 18 March 2021 9 - 1 pm and Session 2 - 25th March 2021 9 - 1pm

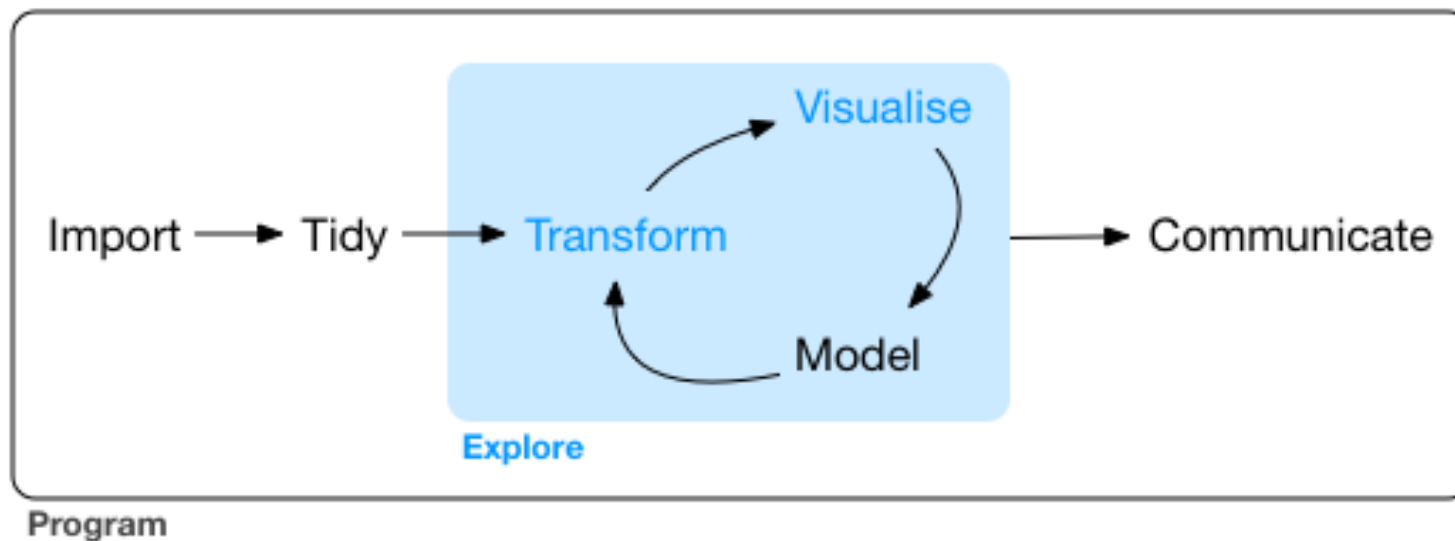
Topic : Exploratory Data Analysis with aimsir17

Course Overview

| Topic | Description |
|------------------|--|
| Session 1 | |
| 1 | <i>Introduction to R and R Studio Cloud</i> |
| 2 | <i>Exploratory Data Analysis: the tibble and ggplot2</i> |
| 3 | <i>Functions, Vectors and Lists</i> |
| 4 | <i>Introduction to Functionals with purrr</i> |
| 5 | <i>Data Transformation I with dplyr</i> |
| Session 2 | |
| 6 | <i>Data Transformation II with dplyr</i> |
| 7 | <i>Statistical Transformation with ggplot2</i> |
| 8 | <i>Functionals and Modelling with purrr</i> |
| 9 | Exploratory Data Analysis - Case Study using aimsir17 |

Exploratory Data Analysis

“Data exploration is the art of looking at your data, rapidly generating hypotheses, quickly testing them, then repeating again and again and again.” (Wickham and Grolemund 2017).



Data Set (1) - observations

```
> observations
```

```
# A tibble: 219,000 x 12
```

| | station | year | month | day | hour | date | rain | temp | rhum | msl | wdsp | wddir |
|----|---------|-------|-------|-------|-------|---------------------|-------|-------|-------|-------|-------|-------|
| | <chr> | <dbl> | <dbl> | <int> | <int> | <dtm> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | ATHENRY | 2017 | 1 | 1 | 0 | 2017-01-01 00:00:00 | 0 | 5.2 | 89 | 1022. | 8 | 320 |
| 2 | ATHENRY | 2017 | 1 | 1 | 1 | 2017-01-01 01:00:00 | 0 | 4.7 | 89 | 1022 | 9 | 320 |
| 3 | ATHENRY | 2017 | 1 | 1 | 2 | 2017-01-01 02:00:00 | 0 | 4.2 | 90 | 1022. | 8 | 320 |
| 4 | ATHENRY | 2017 | 1 | 1 | 3 | 2017-01-01 03:00:00 | 0.1 | 3.5 | 87 | 1022. | 9 | 330 |
| 5 | ATHENRY | 2017 | 1 | 1 | 4 | 2017-01-01 04:00:00 | 0.1 | 3.2 | 89 | 1023. | 8 | 330 |
| 6 | ATHENRY | 2017 | 1 | 1 | 5 | 2017-01-01 05:00:00 | 0 | 2.1 | 91 | 1023. | 8 | 330 |
| 7 | ATHENRY | 2017 | 1 | 1 | 6 | 2017-01-01 06:00:00 | 0 | 2 | 89 | 1024. | 7 | 330 |
| 8 | ATHENRY | 2017 | 1 | 1 | 7 | 2017-01-01 07:00:00 | 0 | 1.7 | 89 | 1024. | 7 | 340 |
| 9 | ATHENRY | 2017 | 1 | 1 | 8 | 2017-01-01 08:00:00 | 0 | 1 | 91 | 1025 | 7 | 330 |
| 10 | ATHENRY | 2017 | 1 | 1 | 9 | 2017-01-01 09:00:00 | 0 | 1.1 | 91 | 1026. | 8 | 330 |

```
# ... with 218,990 more rows
```

Data Set (2) - stations

```
> stations
# A tibble: 25 x 5
  station      county height latitude longitude
  <chr>      <chr>    <dbl>   <dbl>   <dbl>
1 ATHENRY    Galway      40     53.3   -8.79
2 BALLYHAISE Cavan       78     54.1   -7.31
3 BELMULLET  Mayo        9     54.2  -10.0
4 CASEMENT   Dublin      91     53.3   -6.44
5 CLAREMORRIS Mayo       68     53.7   -8.99
6 CORK AIRPORT Cork     155     51.8   -8.49
7 DUBLIN AIRPORT Dublin     71     53.4   -6.24
8 DUNSANY     Meath       83     53.5   -6.66
9 FINNER      Donegal     33     54.5   -8.24
10 GURTEEN    Tipperary   75     53.1   -8.01
# ... with 15 more rows
```

Data Set (3) – eirgrid17

```
> eirgrid17
```

```
# A tibble: 35,040 x 15
```

| | year | month | day | hour | minute | date | NIGeneration | NIDemand | NIWindAvailabil... | NIWindGeneration |
|----|-------|-------|-------|-------|--------|---------------------|--------------|----------|--------------------|------------------|
| | <dbl> | <dbl> | <int> | <int> | <int> | <dtm> | <dbl> | <dbl> | <dbl> | <dbl> |
| 1 | 2017 | 1 | 1 | 0 | 0 | 2017-01-01 00:00:00 | 889. | 776. | 175. | 198. |
| 2 | 2017 | 1 | 1 | 0 | 15 | 2017-01-01 00:15:00 | 922. | 770. | 183. | 208. |
| 3 | 2017 | 1 | 1 | 0 | 30 | 2017-01-01 00:30:00 | 908. | 761. | 170. | 193. |
| 4 | 2017 | 1 | 1 | 0 | 45 | 2017-01-01 00:45:00 | 919. | 743. | 168. | 191. |
| 5 | 2017 | 1 | 1 | 1 | 0 | 2017-01-01 01:00:00 | 882. | 749. | 174. | 196. |
| 6 | 2017 | 1 | 1 | 1 | 15 | 2017-01-01 01:15:00 | 849. | 742. | 190. | 213. |
| 7 | 2017 | 1 | 1 | 1 | 30 | 2017-01-01 01:30:00 | 843. | 726. | 222. | 245. |
| 8 | 2017 | 1 | 1 | 1 | 45 | 2017-01-01 01:45:00 | 809. | 709. | 233. | 245. |
| 9 | 2017 | 1 | 1 | 2 | 0 | 2017-01-01 02:00:00 | 797. | 697. | 282. | 299. |
| 10 | 2017 | 1 | 1 | 2 | 15 | 2017-01-01 02:15:00 | 755. | 684. | 259. | 285. |

```
# ... with 35,030 more rows, and 5 more variables: IEGeneration <dbl>, IEDemand <dbl>, IEWindAvailability <dbl>,
# IEWindGeneration <dbl>, SNSP <chr>
```

Exploratory Data Analysis - Samples

- Weather (Pick 3 stations)
 - West v East v South, is there a difference in annual rainfall, temperature?
- Overall
 - What is the driest month of the year?
 - Association between atmospheric pressure and rainfall?
- Energy
 - Association between average temperature and energy demand?
 - Which weather station's average wind speed produces the best linear model for wind energy generated.