Data Science for Operational Researchers Using R Online

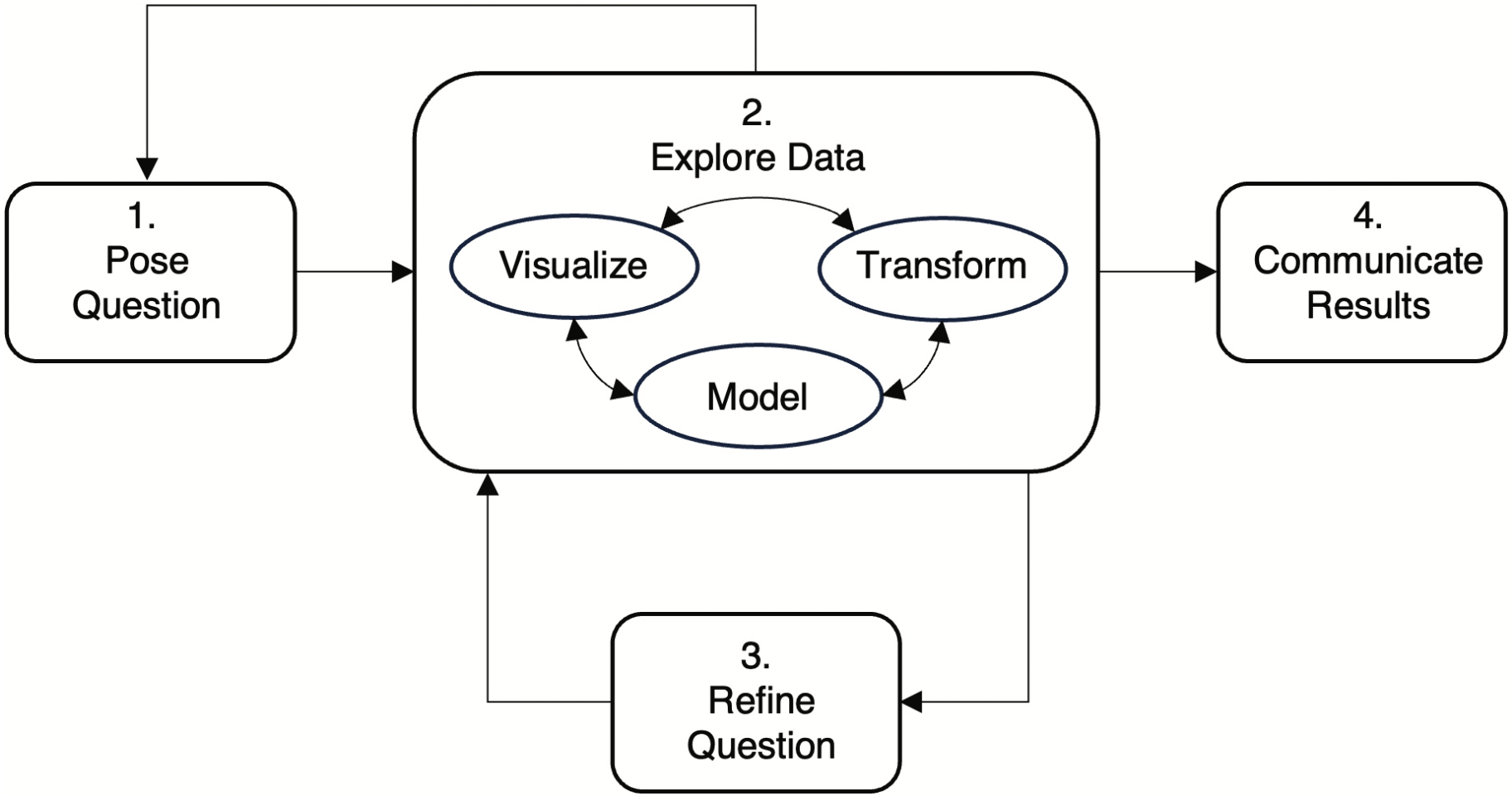
10. Project Work

Prof. Jim Duggan,
School of Computer Science
University of Galway.

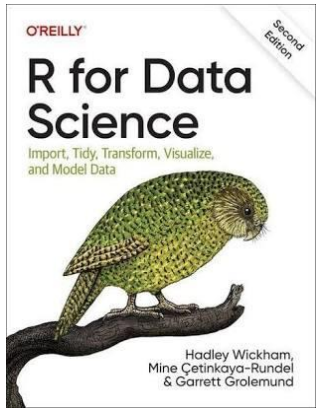
https://github.com/JimDuggan/explore_or

Data packages to explore

1. Overall Process



2. palmerpenguins package (CRAN)



https://r4ds.hadley.nz/data-visualize.html

```
> library(palmerpenguins)
> penguins
# A tibble: 344 × 8
   species island    bill_length_mm bill_depth_mm flipper_length_mm body_mass_g sex     year
   <fct>   <fct>              <dbl>         <dbl>             <int>       <int> <fct>  <int>
 1 Adelie  Torgersen           39.1          18.7               181        3750 male    2007
 2 Adelie  Torgersen           39.5          17.4               186        3800 female  2007
 3 Adelie  Torgersen           40.3          18                 195        3250 female  2007
 4 Adelie  Torgersen           NA            NA                  NA          NA NA      2007
 5 Adelie  Torgersen           36.7          19.3               193        3450 female  2007
 6 Adelie  Torgersen           39.3          20.6               190        3650 male    2007
 7 Adelie  Torgersen           38.9          17.8               181        3625 female  2007
 8 Adelie  Torgersen           39.2          19.6               195        4675 male    2007
 9 Adelie  Torgersen           34.1          18.1               193        3475 NA      2007
10 Adelie  Torgersen           42            20.2               190        4250 NA      2007
# ℹ 334 more rows
# ℹ Use `print(n = ...)` to see more rows
```

2. aimsir17 package (CRAN)

```
> glimpse(stations)
Rows: 25
Columns: 5
$ station   <chr> "ATHENRY",
$ county    <chr> "Galway",
$ height    <dbl> 40, 78, 9,
$ latitude  <dbl> 53.289, 54
$ longitude <dbl> -8.786, -7
```

```
> glimpse(observations)
Rows: 219,000
Columns: 12
$ station <chr> "ATHENRY"
$ year    <dbl> 2017, 201
$ month   <dbl> 1, 1, 1,
$ day     <int> 1, 1, 1,
$ hour    <int> 0, 1, 2,
$ date    <dttm> 2017-01-
$ rain    <dbl> 0.0, 0.0,
$ temp    <dbl> 5.2, 4.7,
$ rhum    <dbl> 89, 89, 9
$ msl     <dbl> 1021.9, 1
$ wdsp    <dbl> 8, 9, 8,
$ wddir   <dbl> 320, 320,
```

```
> glimpse(eirgrid17)
Rows: 35,040
Columns: 15
$ year                 <dbl> 2017,
$ month                <dbl> 1, 1,
$ day                  <int> 1, 1,
$ hour                 <int> 0, 0,
$ minute               <int> 0, 15,
$ date                 <dttm> 2017-
$ NIGeneration         <dbl> 889.00
$ NIDemand             <dbl> 775.93
$ NIWindAvailability   <dbl> 175.06
$ NIWindGeneration     <dbl> 198.20
$ IEGeneration         <dbl> 3288.5
$ IEDemand             <dbl> 2921.4
$ IEWindAvailability   <dbl> 1064.7
$ IEWindGeneration     <dbl> 1044.7
$ SNSP                 <chr> "28.4%
```

3. nycflights23 package (CRAN)

```
> glimpse(weather)
Rows: 26,207
Columns: 15
$ origin     <chr> "JFK"
$ year       <int> 2023,
$ month      <int> 1, 1,
$ day        <int> 1, 1,
$ hour       <int> 0, 1,
$ temp       <dbl> 48.0,
$ dewp       <dbl> 48.0,
$ humid      <dbl> 100.0
$ wind_dir   <dbl> 0, 19
$ wind_speed <dbl> 0.000
$ wind_gust  <dbl> 0.000
$ precip     <dbl> 1e-02
$ pressure   <dbl> 1010.
$ visib      <dbl> 0.25,
$ time_hour  <dttm> 2023
```

```
> glimpse(airlines)
Rows: 14
Columns: 2
$ carrier <chr> "9E", "AA"
$ name    <chr> "Endeavor
```

```
> glimpse(flights)
Rows: 435,352
Columns: 19
$ year           <int> 2023
$ month          <int> 1, 1
$ day            <int> 1, 1
$ dep_time       <int> 1, 1
$ sched_dep_time <int> 2038
$ dep_delay      <dbl> 203,
$ arr_time       <int> 328,
$ sched_arr_time <int> 3, 1
$ arr_delay      <dbl> 205,
$ carrier        <chr> "UA"
$ flight         <int> 628,
$ tailnum        <chr> "N25
$ origin         <chr> "EWR
$ dest           <chr> "SMF
$ air_time       <dbl> 367,
$ distance       <dbl> 2500
$ hour           <dbl> 20,
$ minute         <dbl> 38,
$ time_hour      <dttm> 202
```

```
> glimpse(planes)
Rows: 4,840
Columns: 9
$ tailnum      <chr> "N101DQ", '
$ year         <int> 2020, 2018,
$ type         <chr> "Fixed wing
$ manufacturer <chr> "AIRBUS", '
$ model        <chr> "A321-211",
$ engines      <int> 2, 2, 2, 2,
$ seats        <int> 199, 133, 8
$ speed        <int> 0, 0, 0, 0,
$ engine       <chr> "Turbo-fan'
```

```
> glimpse(airports)
Rows: 1,255
Columns: 8
$ faa   <chr> "AAF", "A
$ name  <chr> "Apalachi
$ lat   <dbl> 29.72750,
$ lon   <dbl> -85.02750
$ alt   <dbl> 20, 79, 3
$ tz    <dbl> -5, -6, -
$ dst   <chr> "A", "A",
$ tzone <chr> "America/
```

4. ISLR2 package (CRAN)

www.statlearning.com

Name	Description
Auto	Gas mileage, horsepower, and other information for cars.
Bikeshare	Hourly usage of a bike sharing program in Washington, DC.
Boston	Housing values and other information about Boston census tracts.
BrainCancer	Survival times for patients diagnosed with brain cancer.
Caravan	Information about individuals offered caravan insurance.
Carseats	Information about car seat sales in 400 stores.
College	Demographic characteristics, tuition, and more for USA colleges.
Credit	Information about credit card debt for 400 customers.
Default	Customer default records for a credit card company.
Fund	Returns of 2,000 hedge fund managers over 50 months.
Hitters	Records and salaries for baseball players.
Khan	Gene expression measurements for four cancer types.
NCI60	Gene expression measurements for 64 cancer cell lines.
NYSE	Returns, volatility, and volume for the New York Stock Exchange.
OJ	Sales information for Citrus Hill and Minute Maid orange juice.
Portfolio	Past values of financial assets, for use in portfolio allocation.
Publication	Time to publication for 244 clinical trials.
Smarket	Daily percentage returns for S&P 500 over a 5-year period.
USArrests	Crime statistics per 100,000 residents in 50 states of USA.
Wage	Income survey data for men in central Atlantic region of USA.
Weekly	1,089 weekly stock market returns for 21 years.