

```
In [ ]: !installed.packages('plyr')
!installed.packages('ggplot2')
!installed.packages('ggExtra')
!installed.packages('tidyverse')
```

```
In [ ]: #library(plyr)
library(ggplot2)
#library("ggExtra")
library(tidyverse)
```

```
— Attaching packages — tidyverse 1.3.0 —
—
```

```
✓ tibble 3.0.3      ✓ dplyr 1.0.2
✓ tidyr 1.1.1      ✓ stringr 1.4.0
✓ readr 1.3.1      ✓ forcats 0.5.0
✓ purrr 0.3.4
```

```
— Conflicts — tidyverse_conflicts() —
—
```

```
✗ dplyr::filter() masks stats::filter()
✗ dplyr::lag()     masks stats::lag()
```

Bài ôn tập: ames_housing

Cho dữ liệu ames_housing.csv. Hãy thực hiện các yêu cầu được liệt kê bên dưới:

1. Đọc dữ liệu, hiển thị thông tin chung của dữ liệu : `head()`, `tail()`, `str()`, `summary()`
2. Cho biết số dòng, số cột của dữ liệu
3. Cho biết có bao nhiêu loại `Garage.Type`, đó là những loại nào, mỗi loại đếm được bao nhiêu mẫu.
4. Xóa các cột `Garage.Qual`, `Garage.Cond`, `Pool.QC`, `Fence`, `Misc.Feature` trong dữ liệu
5. Tìm max, min của `Gr.Liv.Area` theo `Electrical` và `Heating` (sử dụng `group_by()` và `summarize()`).
6. Vẽ biểu đồ thể hiện mối liên hệ của `SalePrice` và `X1st.Flr.SF`. Nhận xét biểu đồ.

7. Vẽ pie chart thể hiện % giữa Y(1) và N(0) của cột Central.Air

8. Cho năm xây của các căn nhà cũ nhất và mới nhất (theo Year.Built). Liệt kê các căn nhà cũ nhất, mới nhất với 3 thông tin Id, Year.Built, SalePrice

9. Thống kê số lượng các căn nhà được xây theo từng năm. In head và tail. Cho biết năm nào có nhiều nhà được xây nhất?

10. Trực quan hóa kết quả của câu thống kê trên với 10 năm gần đây nhất bằng barplot; với tất cả các năm bằng line.

11. Vẽ boxplot của cột SalePrice.

In []: