Data Sets

Ames

```
ames <- AmesHousing::make_ames()

dim(ames)
```

```
[1] 2930   81
```

```
head(ames)
```

```
# A tibble: 6 x 81
  MS_SubClass MS_Zoning Lot_Frontage Lot_Area Street Alley Lot_Shape
  <fct>       <fct>            <dbl>    <int> <fct>  <fct> <fct>
1 One_Story_~ Resident~          141    31770 Pave   No_A~ Slightly~
2 One_Story_~ Resident~           80    11622 Pave   No_A~ Regular
3 One_Story_~ Resident~           81    14267 Pave   No_A~ Slightly~
4 One_Story_~ Resident~           93    11160 Pave   No_A~ Regular
5 Two_Story_~ Resident~           74    13830 Pave   No_A~ Slightly~
6 Two_Story_~ Resident~           78     9978 Pave   No_A~ Slightly~
# ... with 74 more variables: Land_Contour <fct>, Utilities <fct>,
#   Lot_Config <fct>, Land_Slope <fct>, Neighborhood <fct>,
#   Condition_1 <fct>, Condition_2 <fct>, Bldg_Type <fct>,
#   House_Style <fct>, Overall_Qual <fct>, Overall_Cond <fct>,
#   Year_Built <int>, Year_Remod_Add <int>, Roof_Style <fct>,
#   Roof_Matl <fct>, Exterior_1st <fct>, Exterior_2nd <fct>,
#   Mas_Vnr_Type <fct>, Mas_Vnr_Area <dbl>, Exter_Qual <fct>,
#   Exter_Cond <fct>, Foundation <fct>, Bsmt_Qual <fct>, Bsmt_Cond <fct>,
#   Bsmt_Exposure <fct>, BsmtFin_Type_1 <fct>, BsmtFin_SF_1 <dbl>,
#   BsmtFin_Type_2 <fct>, BsmtFin_SF_2 <dbl>, Bsmt_Unf_SF <dbl>,
#   Total_Bsmt_SF <dbl>, Heating <fct>, Heating_QC <fct>,
#   Central_Air <fct>, Electrical <fct>, First_Flr_SF <int>,
#   Second_Flr_SF <int>, Low_Qual_Fin_SF <int>, Gr_Liv_Area <int>,
#   Bsmt_Full_Bath <dbl>, Bsmt_Half_Bath <dbl>, Full_Bath <int>,
#   Half_Bath <int>, Bedroom_AbvGr <int>, Kitchen_AbvGr <int>,
#   Kitchen_Qual <fct>, TotRms_AbvGrd <int>, Functional <fct>,
#   Fireplaces <int>, Fireplace_Qu <fct>, Garage_Type <fct>,
#   Garage_Finish <fct>, Garage_Cars <dbl>, Garage_Area <dbl>,
#   Garage_Qual <fct>, Garage_Cond <fct>, Paved_Drive <fct>,
#   Wood_Deck_SF <int>, Open_Porch_SF <int>, Enclosed_Porch <int>,
#   Three_season_porch <int>, Screen_Porch <int>, Pool_Area <int>,
#   Pool_QC <fct>, Fence <fct>, Misc_Feature <fct>, Misc_Val <int>,
```

```
#   Mo_Sold <int>, Year_Sold <int>, Sale_Type <fct>, Sale_Condition <fct>,
#   Sale_Price <int>, Longitude <dbl>, Latitude <dbl>
```

Attrition

```
attrition <- rsample::attrition

dim(attrition)
```

```
[1] 1470   31
```

```
head(attrition)
```

```
  Age Attrition    BusinessTravel DailyRate           Department
1  41       Yes     Travel_Rarely      1102                Sales
2  49        No Travel_Frequently       279 Research_Development
4  37       Yes     Travel_Rarely      1373 Research_Development
5  33        No Travel_Frequently      1392 Research_Development
7  27        No     Travel_Rarely       591 Research_Development
8  32        No Travel_Frequently      1005 Research_Development
  DistanceFromHome     Education EducationField EnvironmentSatisfaction
1                1       College  Life_Sciences                  Medium
2                8 Below_College  Life_Sciences                    High
4                2       College          Other               Very_High
5                3        Master  Life_Sciences               Very_High
7                2 Below_College        Medical                     Low
8                2       College  Life_Sciences               Very_High
  Gender HourlyRate JobInvolvement JobLevel              JobRole
1 Female         94           High        2      Sales_Executive
2   Male         61         Medium        2   Research_Scientist
4   Male         92         Medium        1 Laboratory_Technician
5 Female         56           High        1   Research_Scientist
7   Male         40           High        1 Laboratory_Technician
8   Male         79           High        1 Laboratory_Technician
  JobSatisfaction MaritalStatus MonthlyIncome MonthlyRate
1       Very_High        Single          5993       19479
2          Medium       Married          5130       24907
4            High        Single          2090        2396
5            High       Married          2909       23159
7          Medium       Married          3468       16632
8       Very_High        Single          3068       11864
  NumCompaniesWorked OverTime PercentSalaryHike PerformanceRating
```

```
1                   8      Yes                11         Excellent
2                   1       No                23       Outstanding
4                   6      Yes                15         Excellent
5                   1      Yes                11         Excellent
7                   9       No                12         Excellent
8                   0       No                13         Excellent
  RelationshipSatisfaction StockOptionLevel TotalWorkingYears
1                      Low                0                 8
2                Very_High                1                10
4                   Medium                0                 7
5                     High                0                 8
7                Very_High                1                 6
8                     High                0                 8
  TrainingTimesLastYear WorkLifeBalance YearsAtCompany YearsInCurrentRole
1                     0             Bad              6                  4
2                     3          Better             10                  7
4                     3          Better              0                  0
5                     3          Better              8                  7
7                     3          Better              2                  2
8                     2            Good              7                  7
  YearsSinceLastPromotion YearsWithCurrManager
1                       0                    5
2                       1                    7
4                       0                    0
5                       3                    0
7                       2                    2
8                       3                    6
```

MNIST

```
mnist <- dslabs::read_mnist()

dim(mnist$train$images)
```

```
[1] 60000   784
```

```
head(mnist$train$labels)
```

```
[1] 5 0 4 1 9 2
```

Grocery

```
url <- "https://koalaverse.github.io/homlr/data/my_basket.csv"

my_basket <- readr::read_csv(url)
```

```
Parsed with column specification:
cols(
  .default = col_double()
)

See spec(...) for full column specifications.
```

```
dim(my_basket)
```

```
[1] 2000   42
```

```
head(my_basket)
```

```
# A tibble: 6 x 42
  `7up` lasagna pepsi   yop red.wine cheese   bbq bulmers mayonnaise
  <dbl>   <dbl> <dbl> <dbl>    <dbl>  <dbl> <dbl>   <dbl>      <dbl>
1     0       0     0     0        0      0     0       0          0
2     0       0     0     0        0      0     0       0          0
3     0       0     0     0        0      0     0       0          0
4     0       0     0     2        1      0     0       0          0
5     0       0     0     0        0      0     0       2          0
6     0       0     0     0        0      0     0       0          0
# ... with 33 more variables: horlics <dbl>, chicken.tikka <dbl>,
#   milk <dbl>, mars <dbl>, coke <dbl>, lottery <dbl>, bread <dbl>,
#   pizza <dbl>, sunny.delight <dbl>, ham <dbl>, lettuce <dbl>,
#   kronenbourg <dbl>, leeks <dbl>, fanta <dbl>, tea <dbl>, whiskey <dbl>,
#   peas <dbl>, newspaper <dbl>, muesli <dbl>, white.wine <dbl>,
#   carrots <dbl>, spinach <dbl>, pate <dbl>, instant.coffee <dbl>,
#   twix <dbl>, potatoes <dbl>, fosters <dbl>, soup <dbl>,
#   toad.in.hole <dbl>, coco.pops <dbl>, kitkat <dbl>, broccoli <dbl>,
#   cigarettes <dbl>
```