Scoring model exercise: Instructions

Exercise 1: Preparations

Optionally, you can use an R Notebook to work on the exercise tasks

1. Set your working directory.
2. Install (if necessary) and load the following packages: `data.table`, `lubridate`, `ggplot2`, and `Hmisc`.
3. Load the data `transactions.csv` with `fread()`.
4. Transpose the variable `TransDate` to datetime with `lubridate()`.

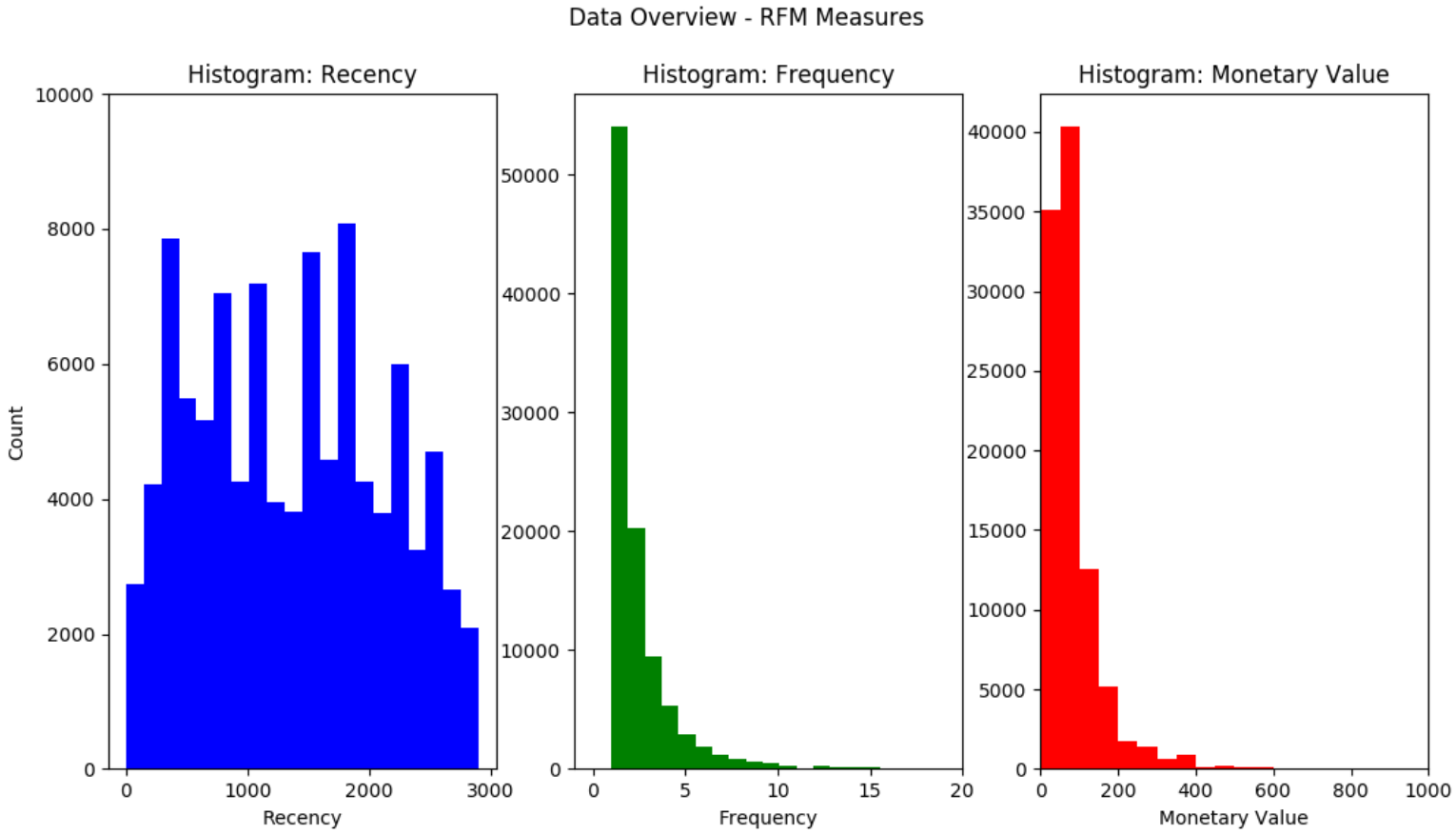
Exercise 2: Aggregation of variables

1. Save the date of the overall latest transaction as the object `now` in your R environment.
2. Create a new data.table called `rfm` that includes the customer ID, as well as the measures for purchase recency, frequency, and monetary value.
3. Check the structure of the new table and ensure that all the variables are numeric.

Exercise 3: Descriptive statistics

1. Get a first impression of the distribution of RFM measures by taking a look at the data summary.
2. Use `ggplot2` in order to plot histograms for all the three measures and arrange them in one single picture.
3. Adjust the title, labels, and colors of your plots in an appealing way.

Exercise 3: Descriptive statistics



Remark: `cut2()` function (1/2)

- **Package:** `Hmisc`
- **Goal:** divide the range of `X` into intervals and code the values in `X` according to which interval they fall.
- **Input:**
  - `X`: numeric vector
  - `g`: number of quantile groups
  - `m`: cuts or observations
- **Output:** Factor object with levels of the form `[a,b]`.

Remark: `cut2()` function (1/2)

- **Package:** `Hmisc`
- **Goal:** divide the range of `X` into intervals and code the values in `X` according to which interval they fall.
- **Input:**
  - `X`: numeric vector
  - `g`: number of quantile groups
  - `m`: cuts or observations
- **Output:** Factor object with levels of the form `[a,b]`.

Remark: `cut2()` function (1/2)

- **Package:** `Hmisc`
- **Goal:** divide the range of `X` into intervals and code the values in `X` according to which interval they fall.
- **Input:**
  - `X`: numeric vector
  - `g`: number of quantile groups
  - `m`: cuts or observations
- **Output:** Factor object with levels of the form `[a,b]`.

Remark: `cut2()` function (2/2)

Rank the values in the vector and divide them in `g=2` equally sized groups (1)

```
temp
```

```
1     1     1     3     4     7     8     9     10    10
```

```
cut2(temp, g=2)
```

```
[1,7]  [1,7]  [1,7]  [1,7]  [1,7]  [7,10] [7,10] [7,10] [7,10] [7,10]
Levels: [1,7] [7,10]
```

```
as.numeric(cut2(temp, g=2))
```

```
1     1     1     1     1     2     2     2     2     2
```

As numeric applied to a factor gives you a vector of the levels the values belong to (2)

Remark: `cut2()` function (2/2)

Rank the values in the vector and divide them in `g=2` equally sized groups (1)

```
temp
```

```
1     1     1     3     4     7     8     9     10    10
```

```
cut2(temp, g=2)
```

```
[1,7]  [1,7]  [1,7]  [1,7]  [1,7]  [7,10] [7,10] [7,10] [7,10] [7,10]
Levels: [1,7] [7,10]
```

```
as.numeric(cut2(temp, g=2))
```

```
1     1     1     1     1     2     2     2     2     2
```

As numeric applied to a factor gives you a vector of the levels the values belong to (2)

Remark: `cut2()` function (2/2)

Rank the values in the vector and divide them in `g=2` equally sized groups (1)

```
temp
```

```
1     1     1     3     4     7     8     9     10    10
```

```
cut2(temp, g=2)
```

```
[1,7]  [1,7]  [1,7]  [1,7]  [1,7]  [7,10]  [7,10]  [7,10]  [7,10]  [7,10]
Levels: [1,7] [7,10]
```

```
as.numeric(cut2(temp, g=2))
```

```
1     1     1     1     1     2     2     2     2     2
```

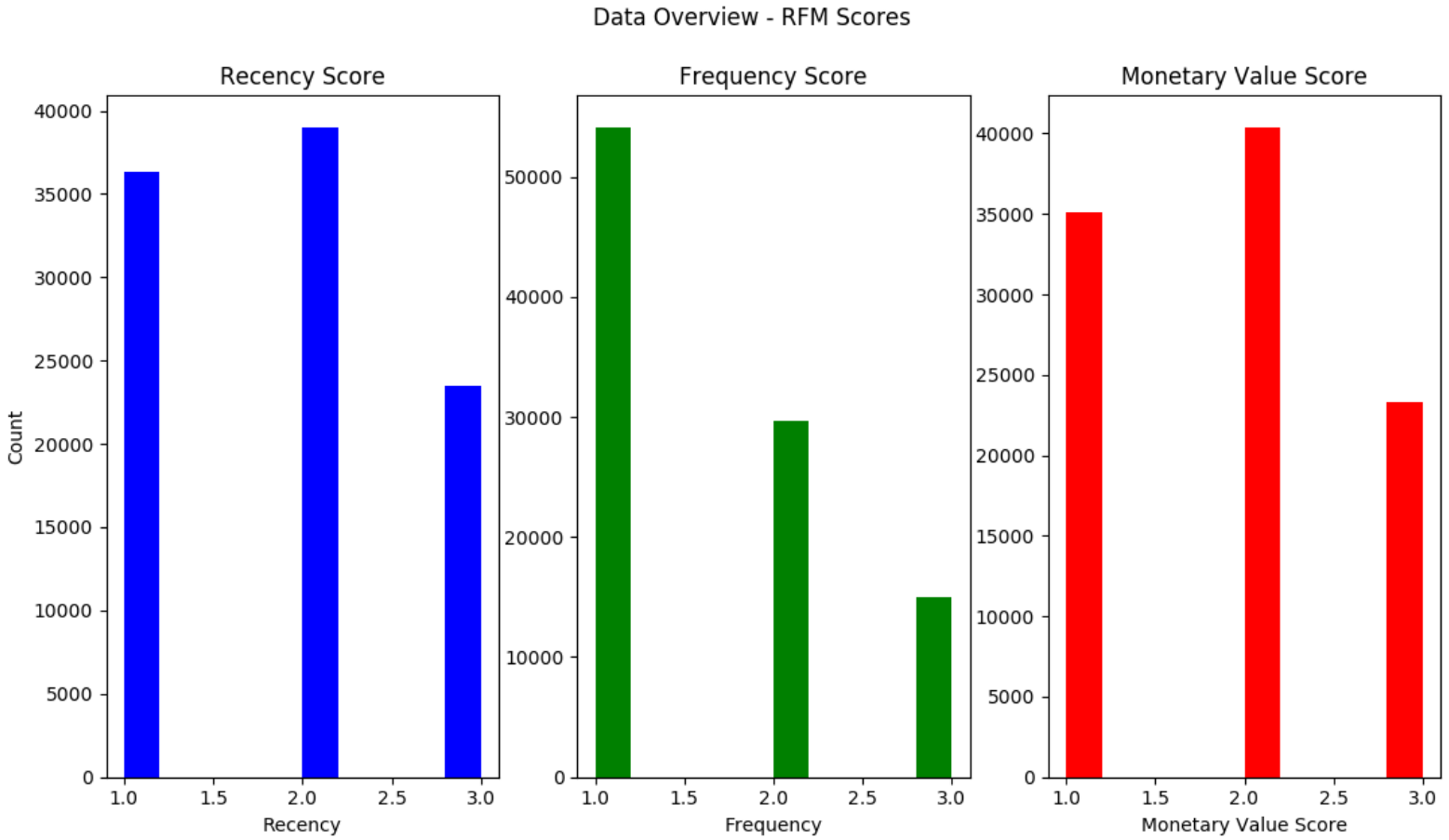
As numeric applied to a factor gives you a vector of the levels the values belong to (2)

Exercise 4: Calculation of RFM scores

1. Use the `cut2()` command to transform the
    1. recency,
    2. frequency, and
    3. monetary value

    measure into scores on a scale from 1 to 3 (develop meaningful category splits). Be aware, that the recency score is decreasing in the days since last purchase!
2. Use `ggplot2` again to plot histograms for all the three scores and arrange them in one single picture. Do you notice something particular?

Exercise 4: Calculation of RFM scores



Exercise 5: The overall RFM score

1. Calculate an unweighted RFM score.
2. Calculate two weighted RFM scores with the following weights: 20-60-20, 60-20-20.
3. Divide all customers into three RFM groups by rounding the unweighted RFM score to the nearest whole number.
4. Display all customers with the highest achievable rating.

Exercise 6: The RFM function

1. Write a function that automates the estimation of the RFM score.
    - **Function input:** data.table as well as the weight for recency, frequency, and monetary value.
    - **Output table:** Customer ID, single RFM scores, overall score, and the RFM group.
2. Ensure that the final RFM score is between 1 and 3 even if the weights do not add up to one.