PART-3: Data transformation

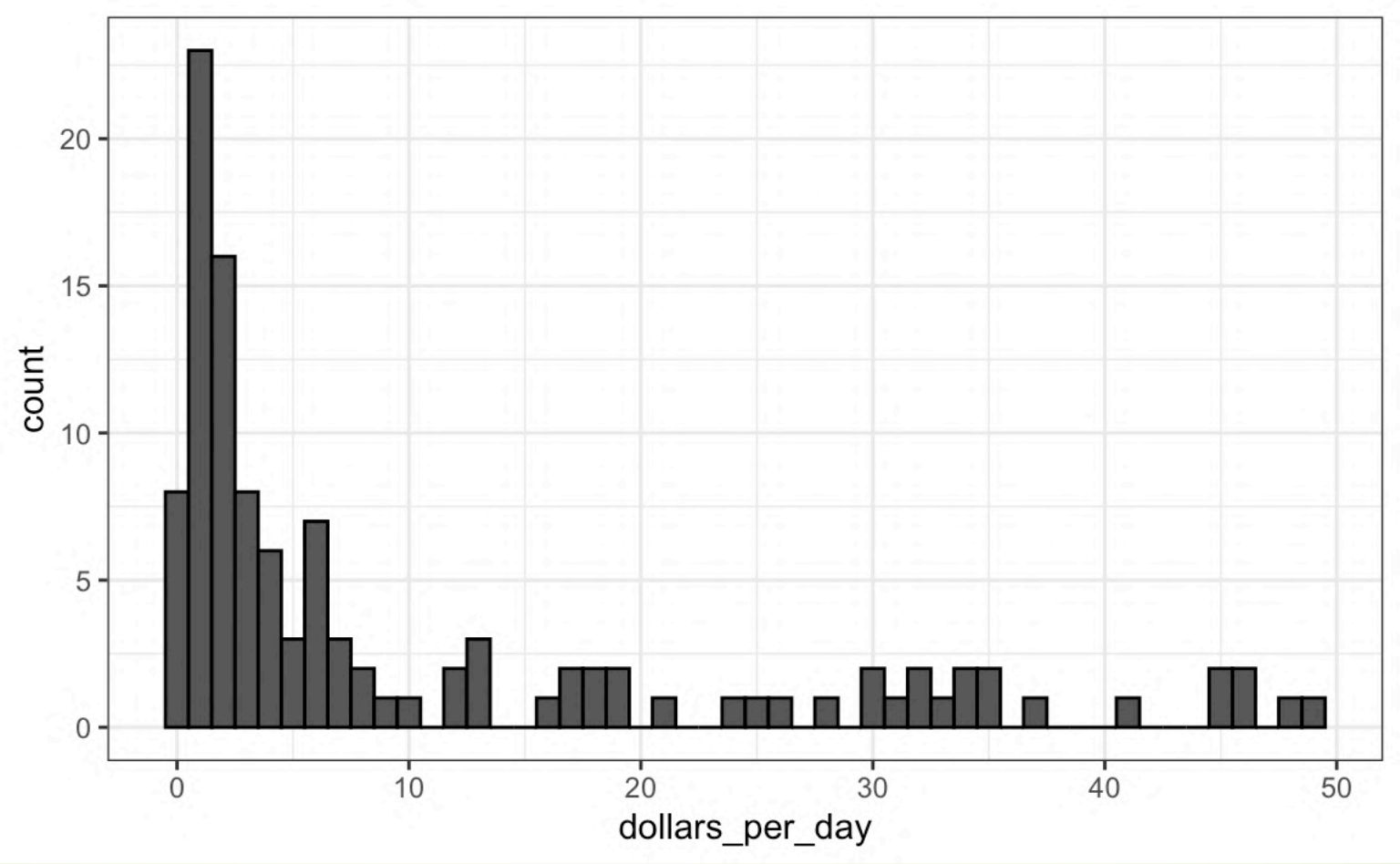
1. Log-transform

1. Log-transform

- Example: Histogram of US per-day income in 1970

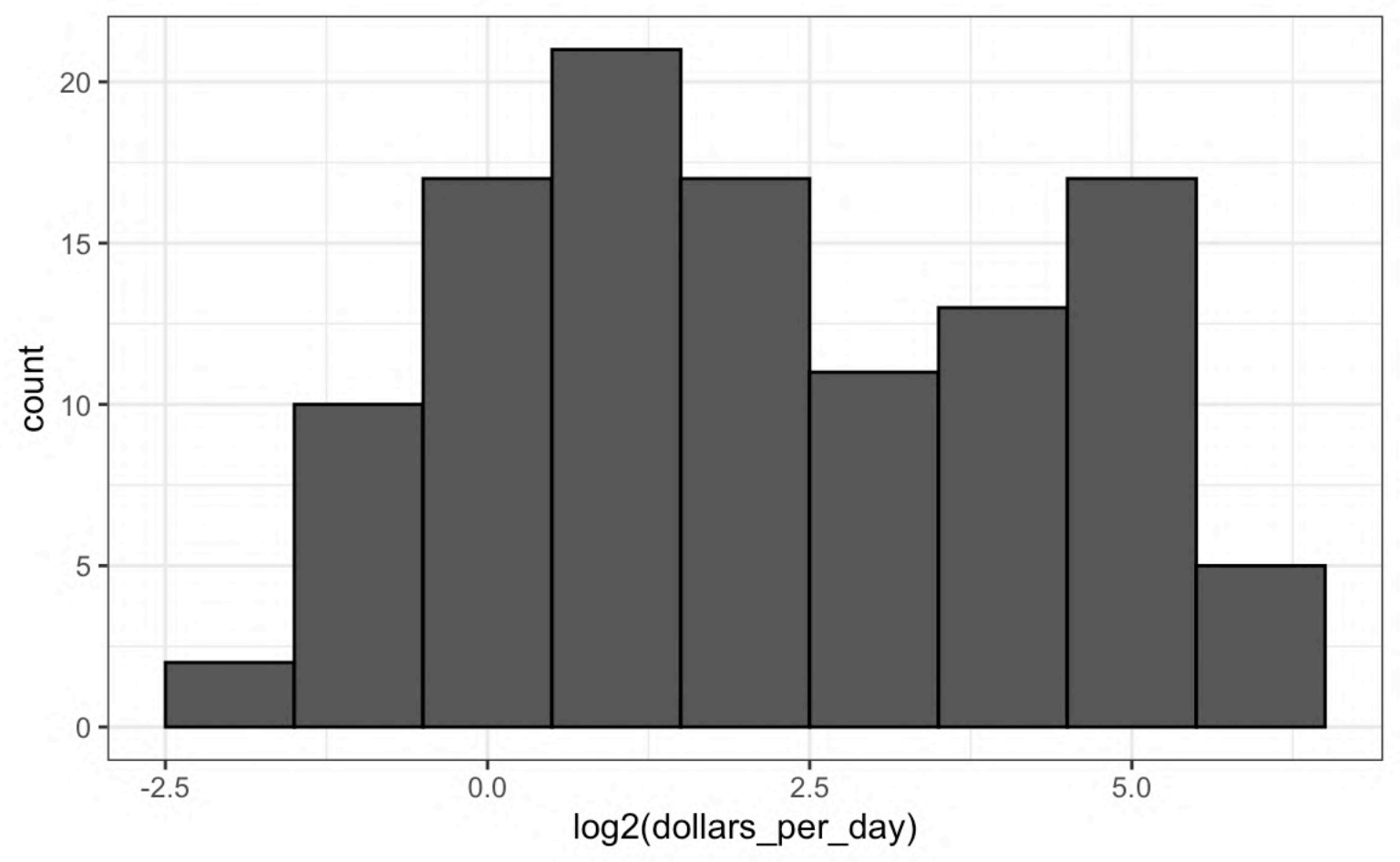
1. Log-transform

- Example: Histogram of US per-day income in 1970



1. Log-transform

- Example: Histogram of US per-day income in 1970



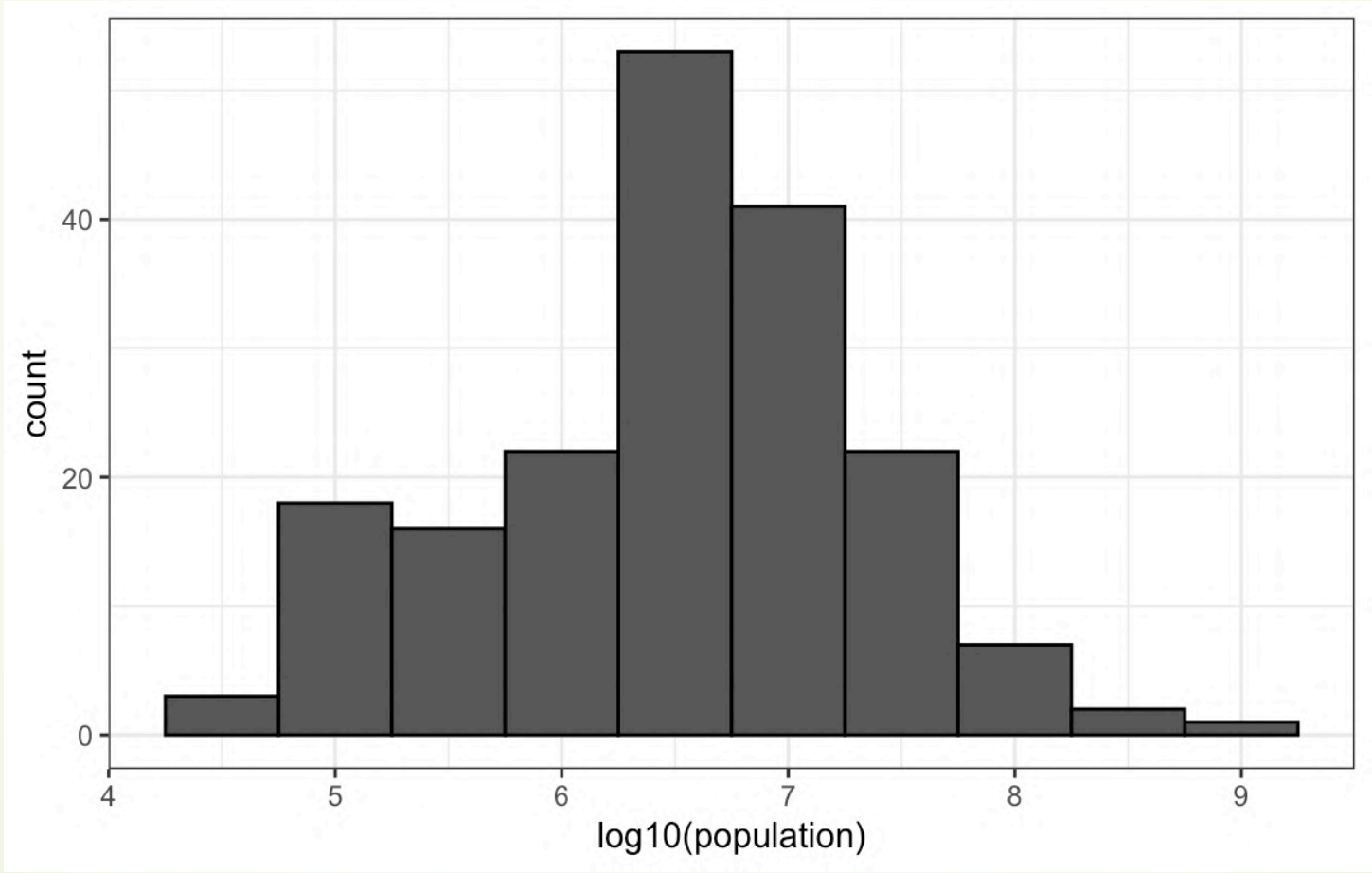
Log-transform

Log-transform

- A different base might be better: Population

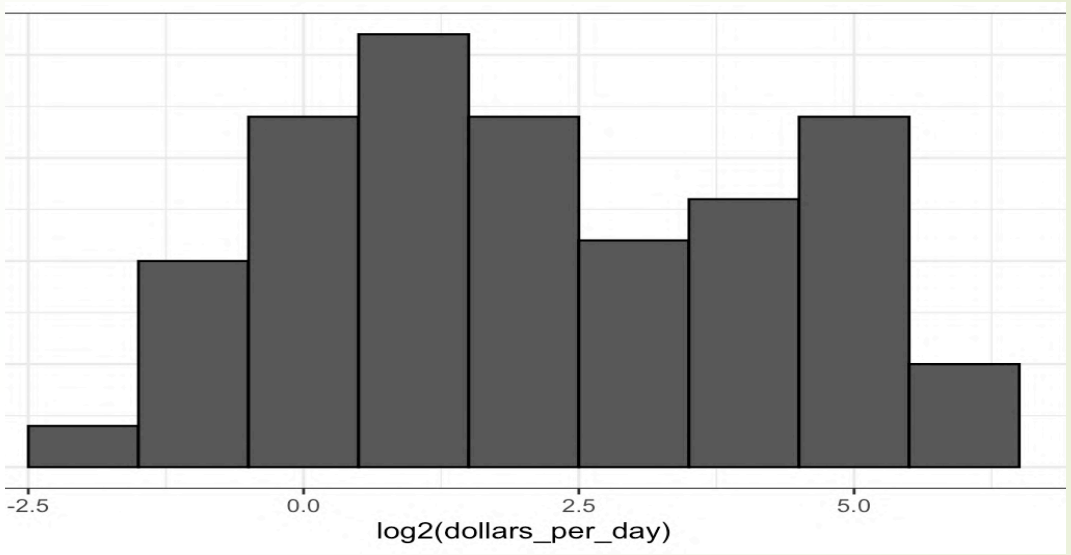
Log-transform

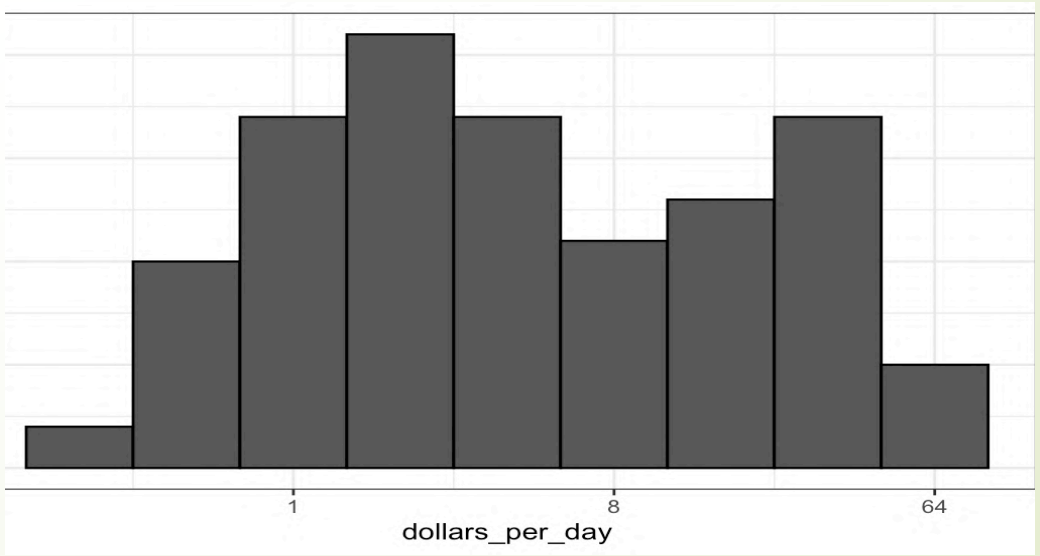
- A different base might be better: Population



Transform value or scale?

- Transforming scale sometimes provides better interpretation





2. Tidying data

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example
- Case study: billboard data

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example
- Case study: billboard data

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example
- Case study: billboard data

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example
- Case study: billboard data

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example
- Case study: billboard data

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example
- Case study: billboard data

2. Tidying data

- Tidying: **Structuring datasets to facilitate analysis**

- Rules of thumb
  - Functional relationships are easiest to see through columns
  - Group comparisons are easiest to see through rows
  - An observation is the smallest unit you'd like to draw conclusions or make predictions about

- *Codd's 3rd normal form*:
  - Each variable forms a column.
  - Each observation forms a row.
  - Each type of observational unit forms a table.

- A simple example
- Case study: billboard data