

# Use of Databases in Data Validation

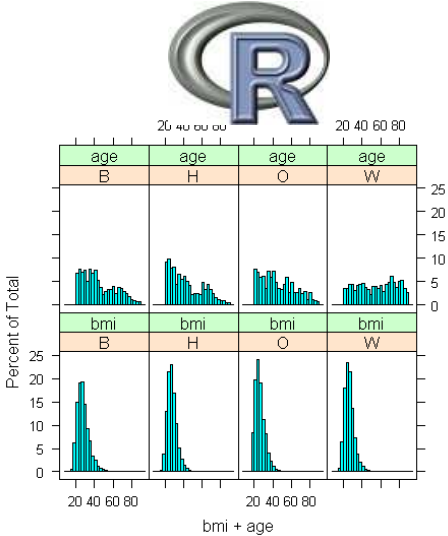
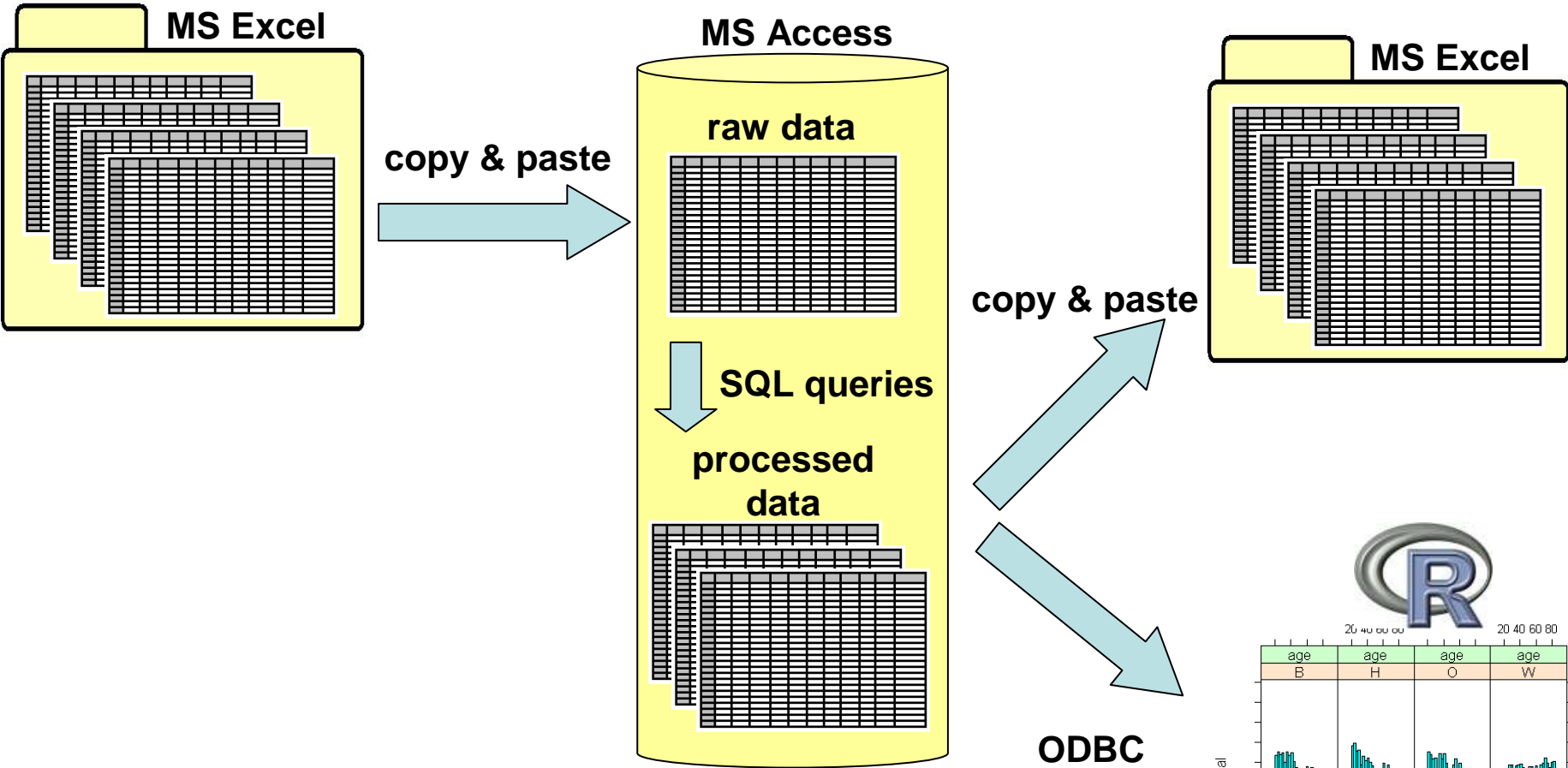
Improving and maintaining  
passport data quality

T. Metz, IRRI

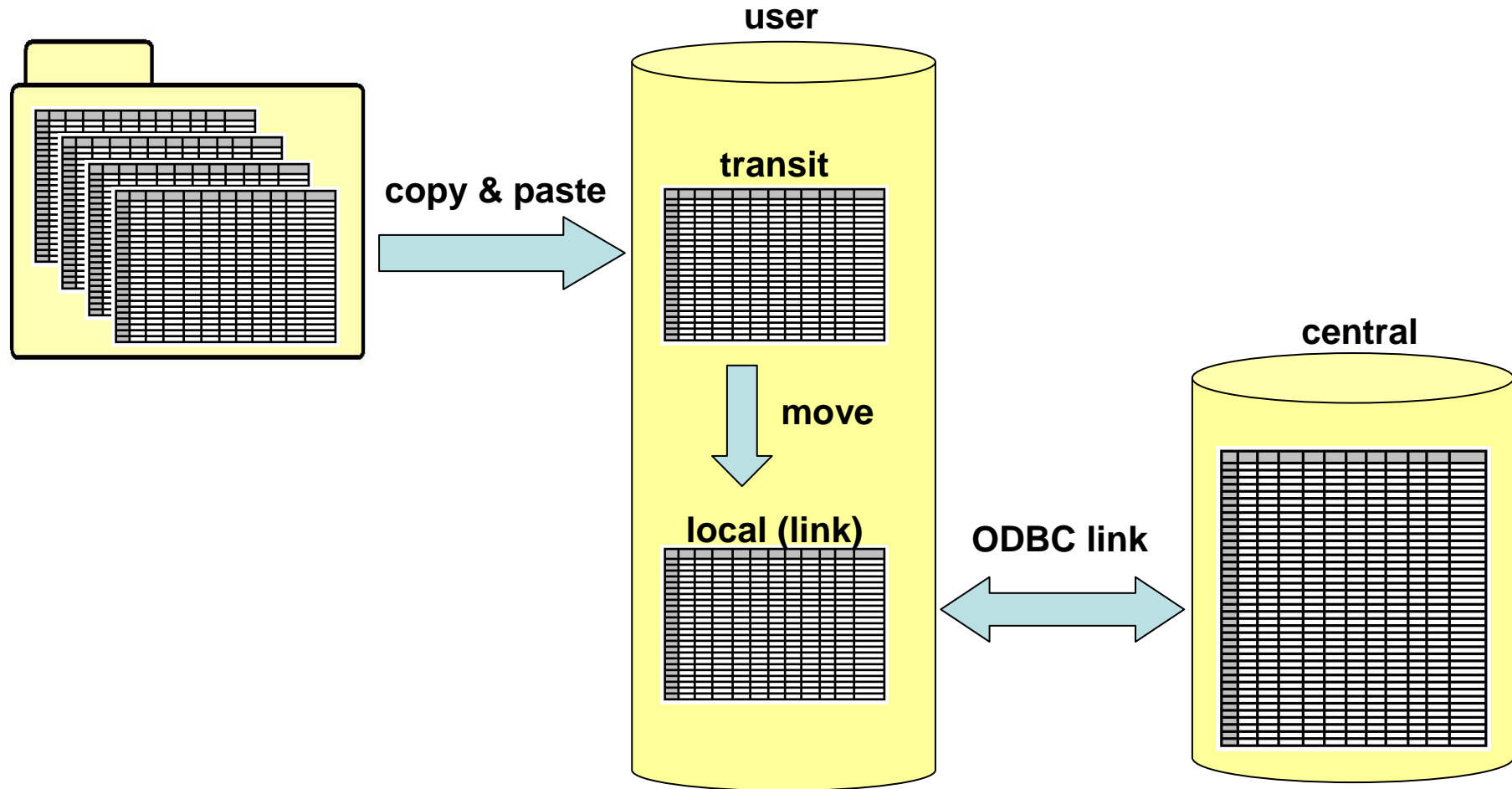
# Data Flow, Treatment, Storage and Use



# Desktop Database



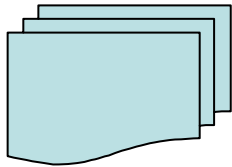
# Desktop & Server Database



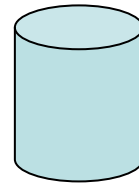
|                       |                           |                      |
|-----------------------|---------------------------|----------------------|
| MS Excel spreadsheets | MS Access database tables | PostgreSQL RDBMS     |
| User desktop / laptop |                           | Institutional server |

# Desktop & Server Database

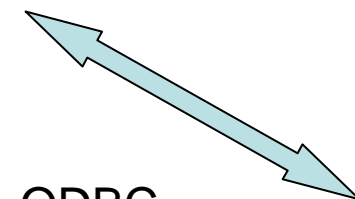
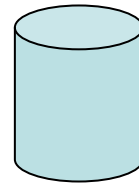
Excel spreadsheets  
conforming to  
database tables



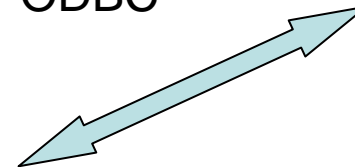
MS Access  
User front-end



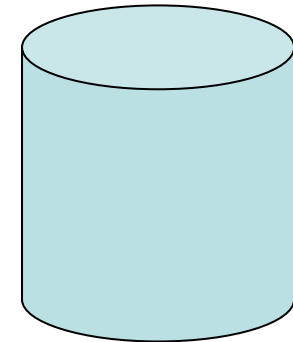
copy & paste



ODBC



Institutional  
Databases  
PostgreSQL



Data capture  
Spreadsheet  
functionality

Linked tables  
Queries  
Reports  
Forms

Schemas  
Views  
Permissions

# The power of databases

- Persistence of data structures
- Strong typing
- Generic NULL values
- Normalization & ease of denormalization
- Powerful queries with SQL
- Access control
- Concurrency control
- Constraints (keys, checks, rules)
- External access
- Near unlimited storage capacity

# Limitations of databases

- Procedural processing (condition, loop, recursion)
- Visualization
- Statistical analysis
- Special functions (financial, trigonometric)
- Output formatting

# Date values

- Most databases have inbuilt date type
- Flexible, human-readable external representation
- Internal representations allows easy date arithmetic
- Problem: incomplete/inaccurate date values
  - Store as string/number (e.g. 20071200)
  - Lose ease of date arithmetic
  - How to code “spring 1975”?
- Solution: best-guess exact date + accuracy
  - Two fields: date + numeric
  - 2007-12-15,15 ; 1975-02-15, 45

# NULL and ternary logic

- NULL is untyped
- NULL is not a value, but a placeholder
- $x = \text{NULL}$  (always unknown)
- $x \text{ IS NULL}$  /  $x \text{ IS NOT NULL}$
- Outer joins create NULL placeholders
- NULL has well-defined behaviour in SQL

| p AND q |         | p       |       |         |
|---------|---------|---------|-------|---------|
|         |         | True    | False | Unknown |
| q       | True    | True    | False | Unknown |
|         | False   | False   | False | False   |
|         | Unknown | Unknown | False | Unknown |

| p OR q |         | p    |         |         |
|--------|---------|------|---------|---------|
|        |         | True | False   | Unknown |
| q      | True    | True | True    | True    |
|        | False   | True | False   | Unknown |
|        | Unknown | True | Unknown | Unknown |

| p       | NOT p   |
|---------|---------|
| True    | False   |
| False   | True    |
| Unknown | Unknown |

# Normalization and de-normalization

|   | cnt_code | cty_name     | cty_pop |
|---|----------|--------------|---------|
| ▶ | AUS      | Adelaide     | 1092369 |
|   | AUS      | Brisbane     | 1598916 |
|   | AUS      | Canberra     | 347612  |
|   | AUS      | Gold Coast   | 390592  |
|   | AUS      | Melbourne    | 3413894 |
|   | AUS      | Newcastle    | 478189  |
|   | AUS      | Perth        | 1361675 |
|   | AUS      | Sydney       | 4031944 |
|   | NZL      | Auckland     | 1105700 |
|   | NZL      | Christchurch | 342000  |
|   | NZL      | Wellington   | 346500  |

Record: 1 of 11

|   | cnt_code | cnt_name    | cnt_pop  |
|---|----------|-------------|----------|
| ▶ | AUS      | Australia   | 18972350 |
|   | NZL      | New Zealand | 3820749  |

Record: 1 of 2

# Normalization and de-normalization

```
join1 : Select Query
SELECT *
FROM city, country;
```

|   | city.cnt_code | cty_name     | cty_pop | country.cnt_code | cnt_name    | cnt_pop  |
|---|---------------|--------------|---------|------------------|-------------|----------|
| ▶ | NZL           | Christchurch | 342000  | AUS              | Australia   | 18972350 |
|   | NZL           | Christchurch | 342000  | NZL              | New Zealand | 3820749  |
|   | NZL           | Wellington   | 346500  | AUS              | Australia   | 18972350 |
|   | NZL           | Wellington   | 346500  | NZL              | New Zealand | 3820749  |
|   | AUS           | Canberra     | 347612  | AUS              | Australia   | 18972350 |
|   | AUS           | Canberra     | 347612  | NZL              | New Zealand | 3820749  |
|   | AUS           | Gold Coast   | 390592  | AUS              | Australia   | 18972350 |
|   | AUS           | Gold Coast   | 390592  | NZL              | New Zealand | 3820749  |
|   | AUS           | Newcastle    | 478189  | AUS              | Australia   | 18972350 |
|   | AUS           | Newcastle    | 478189  | NZL              | New Zealand | 3820749  |
|   | AUS           | Adelaide     | 1092369 | AUS              | Australia   | 18972350 |
|   | AUS           | Adelaide     | 1092369 | NZL              | New Zealand | 3820749  |
|   | NZL           | Auckland     | 1105700 | AUS              | Australia   | 18972350 |
|   | NZL           | Auckland     | 1105700 | NZL              | New Zealand | 3820749  |
|   | AUS           | Perth        | 1361675 | AUS              | Australia   | 18972350 |
|   | AUS           | Perth        | 1361675 | NZL              | New Zealand | 3820749  |
|   | AUS           | Brisbane     | 1598916 | AUS              | Australia   | 18972350 |
|   | AUS           | Brisbane     | 1598916 | NZL              | New Zealand | 3820749  |
|   | AUS           | Melbourne    | 3413894 | AUS              | Australia   | 18972350 |
|   | AUS           | Melbourne    | 3413894 | NZL              | New Zealand | 3820749  |
|   | AUS           | Sydney       | 4031944 | AUS              | Australia   | 18972350 |
|   | AUS           | Sydney       | 4031944 | NZL              | New Zealand | 3820749  |

Record: 1 of 22

# Normalization and de-normalization

```
join2 : Select Query
SELECT *
FROM city, country
WHERE city.cnt_code = country.cnt_code;
```

|   | city.cnt_code | cty_name     | cty_pop | country.cnt_code | cnt_name    | cnt_pop  |
|---|---------------|--------------|---------|------------------|-------------|----------|
| ▶ | AUS           | Canberra     | 347612  | AUS              | Australia   | 18972350 |
|   | AUS           | Gold Coast   | 390592  | AUS              | Australia   | 18972350 |
|   | AUS           | Newcastle    | 478189  | AUS              | Australia   | 18972350 |
|   | AUS           | Adelaide     | 1092369 | AUS              | Australia   | 18972350 |
|   | AUS           | Perth        | 1361675 | AUS              | Australia   | 18972350 |
|   | AUS           | Brisbane     | 1598916 | AUS              | Australia   | 18972350 |
|   | AUS           | Melbourne    | 3413894 | AUS              | Australia   | 18972350 |
|   | AUS           | Sydney       | 4031944 | AUS              | Australia   | 18972350 |
|   | NZL           | Christchurch | 342000  | NZL              | New Zealand | 3820749  |
|   | NZL           | Wellington   | 346500  | NZL              | New Zealand | 3820749  |
|   | NZL           | Auckland     | 1105700 | NZL              | New Zealand | 3820749  |

Record: 1 of 11

# Normalization and de-normalization

```
join3 : Select Query
SELECT cty_name, cty_pop, cnt_name, cnt_pop
FROM city, country
WHERE city.cnt_code = country.cnt_code;
```

|   | cty_name     | cty_pop | cnt_name    | cnt_pop  |
|---|--------------|---------|-------------|----------|
| ▶ | Canberra     | 347612  | Australia   | 18972350 |
|   | Gold Coast   | 390592  | Australia   | 18972350 |
|   | Newcastle    | 478189  | Australia   | 18972350 |
|   | Adelaide     | 1092369 | Australia   | 18972350 |
|   | Perth        | 1361675 | Australia   | 18972350 |
|   | Brisbane     | 1598916 | Australia   | 18972350 |
|   | Melbourne    | 3413894 | Australia   | 18972350 |
|   | Sydney       | 4031944 | Australia   | 18972350 |
|   | Christchurch | 342000  | New Zealand | 3820749  |
|   | Wellington   | 346500  | New Zealand | 3820749  |
|   | Auckland     | 1105700 | New Zealand | 3820749  |

Record: 1 of 11

Microsoft Excel - nhanes.xls

File Edit View Insert Format Tools Data Window Help Type a question for help

Arial 10 B I U

A28

|    | A                                                                                                                                                                                       | B                                                        | C |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------|---|
| 1  | Variable name                                                                                                                                                                           | Variable description                                     |   |
| 2  | =====                                                                                                                                                                                   |                                                          |   |
| 3  | rseqn                                                                                                                                                                                   | Respondent identification number                         |   |
| 4  | fseqn                                                                                                                                                                                   | Family sequence number                                   |   |
| 5  | ethn                                                                                                                                                                                    | Ethnicity: W = White; B = Black; H = Hispanic; O = Other |   |
| 6  | sex                                                                                                                                                                                     | Sex: M = Male; F = Female                                |   |
| 7  | age                                                                                                                                                                                     | Age at interview (see ageunit)                           |   |
| 8  | ageunit                                                                                                                                                                                 | Age at interview-unit: M = Months; Y = Years             |   |
| 9  | weight                                                                                                                                                                                  | Weight (kg) (2 months and over)                          |   |
| 10 | height                                                                                                                                                                                  | Standing height (cm) (2 years and over)                  |   |
| 11 |                                                                                                                                                                                         |                                                          |   |
| 12 | Source of original data:                                                                                                                                                                |                                                          |   |
| 13 | <a href="http://mreed.umtri.umich.edu/mreed/downloads/anthro/NHANESIII/Data/NHANESIIIData.zip">http://mreed.umtri.umich.edu/mreed/downloads/anthro/NHANESIII/Data/NHANESIIIData.zip</a> |                                                          |   |
| 14 | US National Health and Nutrition Survey                                                                                                                                                 |                                                          |   |
| 15 | <a href="http://www.cdc.gov/nchs/about/major/nhanes/nh3data.htm">http://www.cdc.gov/nchs/about/major/nhanes/nh3data.htm</a>                                                             |                                                          |   |
| 16 |                                                                                                                                                                                         |                                                          |   |
| 17 |                                                                                                                                                                                         |                                                          |   |
| 18 |                                                                                                                                                                                         |                                                          |   |
| 19 |                                                                                                                                                                                         |                                                          |   |
| 20 |                                                                                                                                                                                         |                                                          |   |

datadef / data /

Ready

|    | A     | B     | C    | D   | E   | F       | G      | H      | I |
|----|-------|-------|------|-----|-----|---------|--------|--------|---|
| 1  | rseqn | fseqn | ethn | sex | age | ageunit | weight | height |   |
| 2  | 3     | 3872  | H    | M   | 21  | Y       | 81.6   | 178.8  |   |
| 3  | 4     | 4115  | H    | F   | 32  | Y       | 61.65  | 162.2  |   |
| 4  | 7     | 855   | B    | F   | 3   | Y       | 13.05  | 97.3   |   |
| 5  | 9     | 4064  | W    | F   | 48  | Y       | 68.05  | 156.9  |   |
| 6  | 10    | 5386  | W    | M   | 35  | Y       | 92.5   | 177.4  |   |
| 7  | 11    | 8142  | H    | M   | 48  | Y       | 70.6   | 168.2  |   |
| 8  | 12    | 4160  | W    | F   | 2   | M       | 6.35   |        |   |
| 9  | 13    | 9447  | W    | M   | 1   | Y       | 12.55  |        |   |
| 10 | 14    | 6793  | W    | F   | 16  | Y       | 47.5   | 166.9  |   |
| 11 | 16    | 8226  | H    | M   | 16  | Y       | 62.65  | 167.6  |   |
| 12 | 19    | 5689  | B    | M   | 44  | Y       | 86.2   | 178.3  |   |
| 13 | 20    | 7139  | W    | F   | 10  | Y       | 36.15  | 148    |   |
| 14 | 21    | 9582  | B    | F   | 7   | Y       | 21.05  | 120.6  |   |
| 15 | 23    | 8570  | B    | F   | 5   | Y       | 17.95  | 112.1  |   |
| 16 | 26    | 7412  | W    | F   | 1   | Y       | 10.95  |        |   |
| 17 | 27    | 1051  | H    | M   | 4   | Y       | 26.65  | 112.9  |   |
| 18 | 28    | 10224 | B    | M   | 3   | Y       | 17.7   | 100    |   |
| 19 | 30    | 1254  | H    | F   | 11  | Y       | 46.67  | 156    |   |
| 20 | 32    | 1040  | H    | M   | 5   | M       | 9.45   |        |   |

**nhanes : Table**

|   | Field Name | Data Type | Description |
|---|------------|-----------|-------------|
| ? | rseqn      | Number    |             |
|   | fseqn      | Number    |             |
|   | ethn       | Text      |             |
|   | sex        | Text      |             |
|   | age        | Number    |             |
|   | ageunit    | Text      |             |
|   | weight     | Number    |             |
| ▶ | height     | Number    |             |
|   |            |           |             |
|   |            |           |             |
|   |            |           |             |
|   |            |           |             |


Field Properties

General | **Lookup**

|                 |        |
|-----------------|--------|
| Field Size      | Single |
| Format          | Fixed  |
| Decimal Places  | 1      |
| Input Mask      |        |
| Caption         |        |
| Default Value   | Null   |
| Validation Rule |        |
| Validation Text |        |
| Required        | No     |
| Indexed         | No     |
| Smart Tags      |        |

A field name can be up to 64 characters long, including spaces. Press F1 for help on field names.

| nhanes : Table |       |       |      |     |     |         |        |        |
|----------------|-------|-------|------|-----|-----|---------|--------|--------|
|                | rseqn | fseqn | ethn | sex | age | ageunit | weight | height |
| ▶              | 0     | 0     |      |     |     |         |        |        |

Record:  1 of 1

Microsoft Excel - nhanes.xls

File Edit View Insert Format Tools Data Window Help Type a question for help

Arial 10 B I U

H31312 165.2

|    | A     | B     | C    | D   | E   | F       | G      | H      | I |
|----|-------|-------|------|-----|-----|---------|--------|--------|---|
| 1  | rsegn | fsegn | ethn | sex | age | ageunit | weight | height |   |
| 2  | 3     | 3872  | H    | M   | 21  | Y       | 81.6   | 178.8  |   |
| 3  |       |       |      |     | 32  | Y       | 61.65  | 162.2  |   |
| 4  |       |       |      |     | 3   | Y       | 13.05  | 97.3   |   |
| 5  |       |       |      |     | 48  | Y       | 68.05  | 156.9  |   |
| 6  |       |       |      |     | 35  | Y       | 92.5   | 177.4  |   |
| 7  |       |       |      |     | 48  | Y       | 70.6   | 168.2  |   |
| 8  |       |       |      |     | 2   | M       | 6.35   |        |   |
| 9  |       |       |      |     | 1   | Y       | 12.55  |        |   |
| 10 |       |       |      |     | 16  | Y       | 47.5   | 166.9  |   |
| 11 |       |       |      |     | 16  | Y       | 62.65  | 167.6  |   |
| 12 |       |       |      |     | 44  | Y       | 86.2   | 178.3  |   |
| 13 |       |       |      |     | 10  | Y       | 36.15  | 148    |   |
| 14 |       |       |      |     | 7   | Y       | 21.05  | 120.6  |   |
| 15 |       |       |      |     | 5   | Y       | 17.95  | 112.1  |   |
| 16 |       |       |      |     | 1   | Y       | 10.95  |        |   |
| 17 |       |       |      |     | 4   | Y       | 26.65  | 112.9  |   |
| 18 |       |       |      |     | 3   | Y       | 17.7   | 100    |   |
| 19 |       |       |      |     | 11  | Y       | 46.67  | 156    |   |
| 20 |       |       |      |     | 5   | M       | 9.45   |        |   |

Context Menu:

- Cut
- Copy
- Paste
- Paste Special...
- Insert...
- Delete...
- Clear Contents
- Insert Comment
- Format Cells...
- Pick From Drop-down List...
- Create List...
- Hyperlink...
- Look Up...

Ready Sum=1158894210

nhanes : Table

|  | rseqn | fseqn | ethn | sex | age | ageunit | weight | height |
|--|-------|-------|------|-----|-----|---------|--------|--------|
|  | 0     | 0     |      |     |     |         |        |        |


Record: 1 of 1

Context menu options:

- New Record
- Delete Record
- Cut
- Copy
- Paste
- Row Height...

| rseqn | fseqn | ethn | sex | age | ageunit | weight | height |
|-------|-------|------|-----|-----|---------|--------|--------|
| 3     | 3872  | H    | M   | 21  | Y       | 81.60  | 178    |
| 4     | 4115  | H    | F   | 32  | Y       | 61.65  | 162    |
| 7     | 855   | B    | F   | 3   | Y       | 13.05  | 97     |
| 9     | 4064  | W    | F   | 48  | Y       | 68.05  | 156    |
| 10    | 5386  | W    | M   | 35  | Y       | 92.50  | 177    |
| 11    | 8142  | H    | M   | 48  | Y       | 70.60  | 168    |
| 12    | 4160  | W    | F   | 2   | M       | 6.35   |        |
| 13    | 9447  | W    | M   | 1   | Y       | 12.55  |        |
| 14    | 6793  | W    | F   | 16  | Y       | 47.50  | 166    |
| 16    | 8226  | H    | M   | 16  | Y       | 62.65  | 167    |
| 19    | 5689  | B    | M   | 44  | Y       | 86.20  | 178    |
| 20    | 7139  | W    | F   | 10  | Y       | 26.15  | 148    |
| 21    | 9582  | B    | F   |     |         |        | 120    |
| 23    | 8570  | B    | F   |     |         |        | 112    |
| 26    | 7412  | W    | F   |     |         |        | 112    |
| 27    | 1051  | H    | M   |     |         |        | 100    |
| 28    | 10224 | B    | M   |     |         |        | 156    |
| 30    | 1254  | H    | F   |     |         |        |        |
| 32    | 1040  | H    | M   | 5   | M       | 9.45   |        |
| 33    | 8745  | H    | F   | 14  | Y       | 56.00  | 164    |
| 34    | 6144  | B    | F   | 42  | Y       | 57.20  | 159    |
| 35    | 3395  | W    | M   | 6   | Y       | 17.85  | 113    |
| 36    | 2304  | W    | M   | 1   | Y       | 10.30  |        |

**Microsoft Office Access**

 **You are about to paste 31311 record(s).**

Are you sure you want to paste these records?

Record: 1 of 31311

| nhanes : Table |       |       |      |     |     |         |        |        |
|----------------|-------|-------|------|-----|-----|---------|--------|--------|
|                | rseqn | fseqn | ethn | sex | age | ageunit | weight | height |
| ▶              | 3     | 3872  | H    | M   | 21  | Y       | 81.60  | 178.8  |
|                | 4     | 4115  | H    | F   | 32  | Y       | 61.65  | 162.2  |
|                | 7     | 855   | B    | F   | 3   | Y       | 13.05  | 97.3   |
|                | 9     | 4064  | W    | F   | 48  | Y       | 68.05  | 156.9  |
|                | 10    | 5386  | W    | M   | 35  | Y       | 92.50  | 177.4  |
|                | 11    | 8142  | H    | M   | 48  | Y       | 70.60  | 168.2  |
|                | 12    | 4160  | W    | F   | 2   | M       | 6.35   |        |
|                | 13    | 9447  | W    | M   | 1   | Y       | 12.55  |        |
|                | 14    | 6793  | W    | F   | 16  | Y       | 47.50  | 166.9  |
|                | 16    | 8226  | H    | M   | 16  | Y       | 62.65  | 167.6  |
|                | 19    | 5689  | B    | M   | 44  | Y       | 86.20  | 178.3  |
|                | 20    | 7139  | W    | F   | 10  | Y       | 36.15  | 148.0  |
|                | 21    | 9582  | B    | F   | 7   | Y       | 21.05  | 120.6  |
|                | 23    | 8570  | B    | F   | 5   | Y       | 17.95  | 112.1  |
|                | 26    | 7412  | W    | F   | 1   | Y       | 10.95  |        |
|                | 27    | 1051  | H    | M   | 4   | Y       | 26.65  | 112.9  |
|                | 28    | 10224 | B    | M   | 3   | Y       | 17.70  | 100.0  |
|                | 30    | 1254  | H    | F   | 11  | Y       | 46.67  | 156.0  |
|                | 32    | 1040  | H    | M   | 5   | M       | 9.45   |        |
|                | 33    | 8745  | H    | F   | 14  | Y       | 56.00  | 164.6  |

Record: 1 of 31311

# SQL query language

**SELECT** <list of column names>

**FROM** <table name>

**WHERE** <conditions to select rows>

**ORDER BY** <column names for sorting>

;

# SQL query language

**SELECT** rseqn, ethn, sex, age, weight, height

**FROM** nhanes

**WHERE** sex = "M" AND weight >= 130 AND  
height <= 170

**ORDER BY** ethn, age DESC

;

# SQL query language

```
short_heavy_men : Select Query
SELECT rseqn, ethn, sex, age, weight, height
FROM nhanes
WHERE sex = "M" AND weight >= 130 AND height <= 170
ORDER BY ethn, age DESC
;
```

|   | rseqn | ethn | sex | age | weight | height |
|---|-------|------|-----|-----|--------|--------|
| ▶ | 42184 | B    | M   | 36  | 134.05 | 166.8  |
|   | 46356 | B    | M   | 19  | 130.85 | 166.4  |
|   | 47207 | B    | M   | 14  | 133.05 | 168.2  |
|   | 37535 | H    | M   | 32  | 140.00 | 167.6  |
|   | 48868 | H    | M   | 31  | 150.90 | 169.9  |
|   | 46437 | H    | M   | 24  | 130.80 | 168.1  |
|   | 45052 | W    | M   | 67  | 139.15 | 169.5  |
|   | 21775 | W    | M   | 48  | 130.35 | 167.3  |
|   | 5550  | W    | M   | 47  | 181.25 | 164.1  |
|   | 19357 | W    | M   | 28  | 133.00 | 162.6  |

Record: 1 of 10

# SQL query language

```
SELECT ethn, sex, avg(weight) AS avgwgt
```

```
FROM nhanes
```

```
WHERE age >= 50
```

```
GROUP BY ethn, sex
```

```
;
```

# SQL query language

```
avgweight : Select Query
SELECT ethn, sex, avg(weight) AS avgwgt
FROM nhanes
WHERE age >= 50
GROUP BY ethn, sex
;
```

|   | ethn | sex | avgwgt      |
|---|------|-----|-------------|
| ▶ | B    | F   | 76.80394118 |
|   | B    | M   | 79.99699208 |
|   | H    | F   | 67.62638451 |
|   | H    | M   | 77.30564738 |
|   | O    | F   | 65.67435897 |
|   | O    | M   | 72.79342593 |
|   | W    | F   | 67.72631228 |
|   | W    | M   | 80.13096967 |

Record: ⏪ ⏩ 1 ▶ ▶▶ ▶▶▶

adultbmi : Select Query

```
SELECT rseqn, fseqn, ethn, sex, age, ageunit, weight, height,  
       round(weight/((height/100)*(height/100)),2) AS bmi  
FROM rhanes  
WHERE ageunit = "Y" AND age >= 20;
```

| adultbmi : Select Query |       |       |      |     |     |         |        |        |       |
|-------------------------|-------|-------|------|-----|-----|---------|--------|--------|-------|
|                         | rseqn | fseqn | ethn | sex | age | ageunit | weight | height | bmi   |
| ▶                       | 3     | 3872  | H    | M   | 21  | Y       | 81.60  | 178.8  | 25.52 |
|                         | 4     | 4115  | H    | F   | 32  | Y       | 61.65  | 162.2  | 23.43 |
|                         | 9     | 4064  | W    | F   | 48  | Y       | 68.05  | 156.9  | 27.64 |
|                         | 10    | 5386  | W    | M   | 35  | Y       | 92.50  | 177.4  | 29.39 |
|                         | 11    | 8142  | H    | M   | 48  | Y       | 70.60  | 168.2  | 24.95 |
|                         | 19    | 5689  | B    | M   | 44  | Y       | 86.20  | 178.3  | 27.11 |
|                         | 34    | 6144  | B    | F   | 42  | Y       | 57.20  | 159.0  | 22.63 |
|                         | 44    | 5204  | W    | F   | 24  | Y       | 55.80  | 163.5  | 20.87 |
|                         | 45    | 7753  | B    | F   | 67  | Y       | 68.00  | 163.4  | 25.47 |
|                         | 48    | 6556  | W    | F   | 56  | Y       | 109.05 | 171.6  | 37.03 |
|                         | 49    | 1295  | W    | F   | 82  | Y       | 43.63  | 151.2  | 19.08 |
|                         | 51    | 364   | H    | M   | 44  | Y       | 144.50 | 180.5  | 44.35 |
|                         | 52    | 7966  | H    | M   | 50  | Y       | 79.60  | 178.2  | 25.07 |
|                         | 53    | 5312  | O    | M   | 36  | Y       | 50.80  | 154.9  | 21.17 |
|                         | 55    | 4289  | H    | M   | 48  | Y       | 111.65 | 172.6  | 37.48 |
|                         | 56    | 9331  | B    | M   | 32  | Y       | 84.20  | 186.4  | 24.23 |
|                         | 63    | 9726  | H    | F   | 66  | Y       | 56.90  | 155.3  | 23.59 |
|                         | 67    | 5969  | W    | M   | 70  | Y       | 64.00  | 162.0  | 24.39 |
|                         | 70    | 1673  | B    | M   | 63  | Y       | 92.23  | 172.1  | 31.14 |







Record: 1 of 17030






summary\_bmi : Select Query

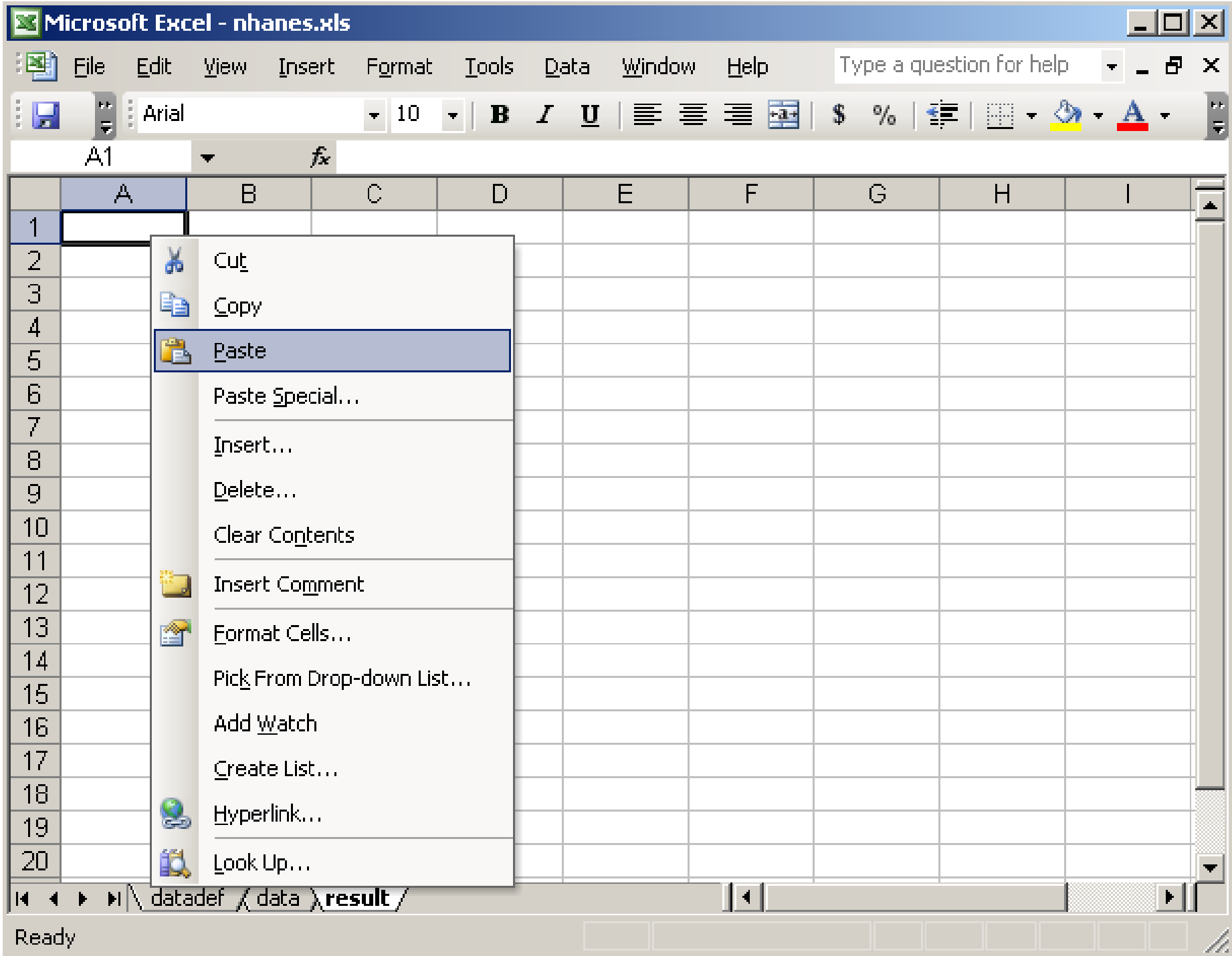
```
SELECT ethn, sex,  
       round(avg(bmi),2) AS avg_bmi,  
       round(stdev(bmi),2) AS std_bmi,  
       min(bmi) AS min_bmi,  
       max(bmi) AS max_bmi,  
       count(bmi) AS cnt_mbi  
FROM adultbmi  
GROUP BY ethn, sex  
ORDER BY ethn, sex;
```

| summary_bmi : Select Query |      |     |         |         |         |         |         |
|----------------------------|------|-----|---------|---------|---------|---------|---------|
|                            | ethn | sex | avg_bmi | std_bmi | min_bmi | max_bmi | cnt_mbi |
| ▶                          | B    | F   | 28.98   | 7.35    | 12.52   | 79.57   | 2580    |
|                            | B    | M   | 26.52   | 5.52    | 12.18   | 70.19   | 2112    |
|                            | H    | F   | 28.05   | 6.1     | 15.46   | 59.66   | 2249    |
|                            | H    | M   | 26.99   | 4.58    | 15.29   | 52.28   | 2250    |
|                            | O    | F   | 26.68   | 5.96    | 16.89   | 50.74   | 382     |
|                            | O    | M   | 25.56   | 4.52    | 17.73   | 51.59   | 286     |
|                            | W    | F   | 26.29   | 6       | 11.71   | 68.54   | 3825    |
|                            | W    | M   | 26.5    | 4.66    | 14.37   | 67.31   | 3285    |

Record: 1 of 8

| summary_bmi : Select Query                                                                      |         |         |         |         |         |  |
|-------------------------------------------------------------------------------------------------|---------|---------|---------|---------|---------|--|
|                                                                                                 | avg_bmi | std_bmi | min_bmi | max_bmi | cnt_mbi |  |
|  New Record    | 28.98   | 7.35    | 12.52   | 79.57   | 2580    |  |
|  Delete Record | 26.52   | 5.52    | 12.18   | 70.19   | 2112    |  |
|  Cut           | 28.05   | 6.1     | 15.46   | 59.66   | 2249    |  |
|  Copy          | 26.99   | 4.58    | 15.29   | 52.28   | 2250    |  |
|  Paste         | 26.68   | 5.96    | 16.89   | 50.74   | 382     |  |
|  Row Height... | 25.56   | 4.52    | 17.73   | 51.59   | 286     |  |
|                                                                                                 | 26.29   | 6       | 11.71   | 68.54   | 3825    |  |
|                                                                                                 | 26.5    | 4.66    | 14.37   | 67.31   | 3285    |  |

Record:   1    of 8



Microsoft Excel - nhanes.xls

File Edit View Insert Format Tools Data Window Help Type a question for help

Arial 10 B I U

A1 & ethn

|    | A    | B   | C       | D       | E       | F       | G       | H | I |
|----|------|-----|---------|---------|---------|---------|---------|---|---|
| 1  | ethn | sex | avg_bmi | std_bmi | min_bmi | max_bmi | cnt_mbi |   |   |
| 2  | B    | F   | 28.98   | 7.35    | 12.52   | 79.57   | 2580    |   |   |
| 3  | B    | M   | 26.52   | 5.52    | 12.18   | 70.19   | 2112    |   |   |
| 4  | H    | F   | 28.05   | 6.1     | 15.46   | 59.66   | 2249    |   |   |
| 5  | H    | M   | 26.99   | 4.58    | 15.29   | 52.28   | 2250    |   |   |
| 6  | O    | F   | 26.68   | 5.96    | 16.89   | 50.74   | 382     |   |   |
| 7  | O    | M   | 25.56   | 4.52    | 17.73   | 51.59   | 286     |   |   |
| 8  | W    | F   | 26.29   | 6       | 11.71   | 68.54   | 3825    |   |   |
| 9  | W    | M   | 26.5    | 4.66    | 14.37   | 67.31   | 3285    |   |   |
| 10 |      |     |         |         |         |         |         |   |   |
| 11 |      |     |         |         |         |         |         |   |   |
| 12 |      |     |         |         |         |         |         |   |   |
| 13 |      |     |         |         |         |         |         |   |   |
| 14 |      |     |         |         |         |         |         |   |   |
| 15 |      |     |         |         |         |         |         |   |   |
| 16 |      |     |         |         |         |         |         |   |   |
| 17 |      |     |         |         |         |         |         |   |   |
| 18 |      |     |         |         |         |         |         |   |   |
| 19 |      |     |         |         |         |         |         |   |   |
| 20 |      |     |         |         |         |         |         |   |   |

datadef data result

Ready Sum=17845.29

| outlier_bmi : Select Query |       |       |      |     |     |         |        |        |       |         |         |     |
|----------------------------|-------|-------|------|-----|-----|---------|--------|--------|-------|---------|---------|-----|
|                            | rseqn | fseqn | ethn | sex | age | ageunit | weight | height | bmi   | avg_bmi | std_bmi | dev |
| ▶                          | 5550  | 2103  | W    | M   | 47  | Y       | 181.25 | 164.1  | 67.31 | 26.5    | 4.66    | 8.8 |
|                            | 44084 | 12975 | B    | M   | 33  | Y       | 241.80 | 185.6  | 70.19 | 26.52   | 5.52    | 7.9 |
|                            | 45394 | 14058 | B    | M   | 23  | Y       | 218.90 | 180.3  | 67.34 | 26.52   | 5.52    | 7.4 |
|                            | 39707 | 18328 | W    | M   | 28  | Y       | 179.70 | 173.7  | 59.56 | 26.5    | 4.66    | 7.1 |
|                            | 38690 | 11115 | W    | F   | 39  | Y       | 175.45 | 160.0  | 68.54 | 26.29   | 6       | 7   |
|                            | 19135 | 7398  | W    | M   | 35  | Y       | 202.47 | 185.5  | 58.84 | 26.5    | 4.66    | 6.9 |
|                            | 43380 | 15267 | B    | F   | 34  | Y       | 213.50 | 163.8  | 79.57 | 28.98   | 7.35    | 6.9 |
|                            | 44490 | 16795 | B    | M   | 30  | Y       | 188.50 | 172.6  | 63.27 | 26.52   | 5.52    | 6.7 |
|                            | 42010 | 16304 | W    | F   | 42  | Y       | 158.60 | 156.8  | 64.51 | 26.29   | 6       | 6.4 |
|                            | 5486  | 9781  | B    | M   | 27  | Y       | 206.44 | 182.4  | 62.05 | 26.52   | 5.52    | 6.4 |
|                            | 1053  | 1750  | W    | M   | 42  | Y       | 178.60 | 178.7  | 55.93 | 26.5    | 4.66    | 6.3 |
|                            | 2350  | 9887  | O    | M   | 41  | Y       | 169.20 | 181.1  | 51.59 | 25.56   | 4.52    | 5.8 |
|                            | 48868 | 13274 | H    | M   | 31  | Y       | 150.90 | 169.9  | 52.28 | 26.99   | 4.58    | 5.5 |
|                            | 51936 | 16327 | H    | F   | 40  | Y       | 150.25 | 158.7  | 59.66 | 28.05   | 6.1     | 5.2 |
|                            | 34424 | 17188 | W    | M   | 41  | Y       | 182.25 | 190.2  | 50.38 | 26.5    | 4.66    | 5.1 |
|                            | 19357 | 7764  | W    | M   | 28  | Y       | 133.00 | 162.6  | 50.3  | 26.5    | 4.66    | 5.1 |
|                            | 20167 | 8428  | B    | M   | 51  | Y       | 168.75 | 176.3  | 54.29 | 26.52   | 5.52    | 5   |
|                            | 51208 | 12948 | H    | F   | 43  | Y       | 132.25 | 150.5  | 58.39 | 28.05   | 6.1     | 5   |
|                            | 37535 | 15805 | H    | M   | 32  | Y       | 140.00 | 167.6  | 49.84 | 26.99   | 4.58    | 5   |
|                            | 9377  | 4941  | H    | F   | 36  | Y       | 138.31 | 154.3  | 58.09 | 28.05   | 6.1     | 4.9 |
|                            | 51038 | 14574 | H    | F   | 35  | Y       | 144.50 | 157.6  | 58.18 | 28.05   | 6.1     | 4.9 |
|                            | 38908 | 12320 | H    | F   | 29  | Y       | 145.35 | 158.8  | 57.64 | 28.05   | 6.1     | 4.9 |
|                            | 43832 | 19535 | W    | M   | 51  | Y       | 151.10 | 174.8  | 49.45 | 26.5    | 4.66    | 4.9 |
|                            | 45052 | 16537 | W    | M   | 67  | Y       | 139.15 | 169.5  | 48.43 | 26.5    | 4.66    | 4.7 |
|                            | 46265 | 16921 | H    | M   | 60  | Y       | 145.35 | 173.5  | 48.29 | 26.99   | 4.58    | 4.7 |

Record: 1 of 208

| bmi_cat_iupui : Table |     |          |        |                          |
|-----------------------|-----|----------|--------|--------------------------|
|                       | sex | from_bmi | to_bmi | status                   |
|                       | M   | 0.0      | 20.7   | underweight              |
|                       | M   | 20.7     | 26.4   | ideal weight             |
|                       | M   | 26.4     | 27.8   | marginally overweight    |
|                       | M   | 27.8     | 31.1   | overweight               |
|                       | M   | 31.1     | 45.5   | very overweight or obese |
|                       | M   | 45.5     | 500.0  | morbid obesity           |
|                       | F   | 0.0      | 19.1   | underweight              |
|                       | F   | 19.1     | 25.8   | ideal weight             |
|                       | F   | 25.8     | 27.3   | marginally overweight    |
|                       | F   | 27.3     | 32.3   | overweight               |
|                       | F   | 32.3     | 44.8   | very overweight or obese |
|                       | F   | 44.8     | 500.0  | morbid obesity           |
| ▶                     |     |          |        |                          |

Record: ⏪ ⏴  ⏵ ⏩ ⏭ of 13

```

categorize2 : Select Query
SELECT adultbmi.rseqn, adultbmi.fseqn, adultbmi.ethn, adultbmi.sex,
       adultbmi.age, adultbmi.ageunit, adultbmi.weight, adultbmi.height,
       adultbmi.bmi, bmi_cat_jupui.status
FROM adultbmi, bmi_cat_jupui
WHERE adultbmi.sex = bmi_cat_jupui.sex AND
       adultbmi.bmi >= bmi_cat_jupui.from_bmi AND
       adultbmi.bmi < bmi_cat_jupui.to_bmi
ORDER BY adultbmi.rseqn;
```

| categorize2 : Select Query |       |       |      |     |     |         |        |        |       |                          |
|----------------------------|-------|-------|------|-----|-----|---------|--------|--------|-------|--------------------------|
|                            | rseqn | fseqn | ethn | sex | age | ageunit | weight | height | bmi   | status                   |
| ▶                          | 3     | 3872  | H    | M   | 21  | Y       | 81.60  | 178.8  | 25.52 | ideal weight             |
|                            | 4     | 4115  | H    | F   | 32  | Y       | 61.65  | 162.2  | 23.43 | ideal weight             |
|                            | 9     | 4064  | W    | F   | 48  | Y       | 68.05  | 156.9  | 27.64 | overweight               |
|                            | 10    | 5386  | W    | M   | 35  | Y       | 92.50  | 177.4  | 29.39 | overweight               |
|                            | 11    | 8142  | H    | M   | 48  | Y       | 70.60  | 168.2  | 24.95 | ideal weight             |
|                            | 19    | 5689  | B    | M   | 44  | Y       | 86.20  | 178.3  | 27.11 | marginally overweight    |
|                            | 34    | 6144  | B    | F   | 42  | Y       | 57.20  | 159.0  | 22.63 | ideal weight             |
|                            | 44    | 5204  | W    | F   | 24  | Y       | 55.80  | 163.5  | 20.87 | ideal weight             |
|                            | 45    | 7753  | B    | F   | 67  | Y       | 68.00  | 163.4  | 25.47 | ideal weight             |
|                            | 48    | 6556  | W    | F   | 56  | Y       | 109.05 | 171.6  | 37.03 | very overweight or obese |
|                            | 49    | 1295  | W    | F   | 82  | Y       | 43.63  | 151.2  | 19.08 | underweight              |
|                            | 51    | 364   | H    | M   | 44  | Y       | 144.50 | 180.5  | 44.35 | very overweight or obese |
|                            | 52    | 7966  | H    | M   | 50  | Y       | 79.60  | 178.2  | 25.07 | ideal weight             |
|                            | 53    | 5312  | O    | M   | 36  | Y       | 50.80  | 154.9  | 21.17 | ideal weight             |
|                            | 55    | 4289  | H    | M   | 48  | Y       | 111.65 | 172.6  | 37.48 | very overweight or obese |
|                            | 56    | 9331  | B    | M   | 32  | Y       | 84.20  | 186.4  | 24.23 | ideal weight             |
|                            | 63    | 9726  | H    | F   | 66  | Y       | 56.90  | 155.3  | 23.59 | ideal weight             |
|                            | 67    | 5969  | W    | M   | 70  | Y       | 64.00  | 162.0  | 24.39 | ideal weight             |
|                            | 70    | 1673  | B    | M   | 63  | Y       | 92.23  | 172.1  | 31.14 | very overweight or obese |
|                            | 71    | 4110  | H    | F   | 37  | Y       | 68.95  | 156.8  | 28.04 | overweight               |

Record: 1 of 16969

# SQL query

```
freq: soil_tex : Select Query
SELECT soil_tex, count(*) AS cnt
FROM passport
GROUP BY soil_tex;
```

| soil_tex   | cnt  |
|------------|------|
|            | 8279 |
| Y          | 1    |
| Clay       | 1757 |
| Clay loam  | 2    |
| Clay loam  | 626  |
| Clay/loam  | 3    |
| Gravel     | 6    |
| Gravel Loa | 1    |
| Gravelly   | 3    |
| Heavy loam | 1    |
| Loam       | 648  |
| Loamy clay | 2    |
| Loamy sand | 129  |
| Organic    | 1    |
| Rock       | 2    |
| Sand       | 120  |
| Sandy      | 659  |
| Sandy clay | 170  |
| Sandy loam | 768  |
| Sandy silt | 1    |
| Savanna    | 5    |
| Silt       | 9    |
| Silt laom  | 1    |
| Silt loam  | 1    |
| Silty loam | 4    |
| Silty sand | 1    |
| U          | 2    |
| Y          | 1    |

Record: 1

# SQL query

```
freq: family-by-pltype : Select Query
SELECT family, pltype, count(*) AS cnt
FROM gensp
GROUP BY family, pltype;
```

|   | family                | pltype | cnt |
|---|-----------------------|--------|-----|
| ▶ | <b>Acanthaceae</b>    | O      | 1   |
|   | <b>Aizoaceae</b>      | B      | 1   |
|   | <b>Amaranthaceae</b>  | B      | 3   |
|   | <b>Amaranthaceae</b>  | O      | 15  |
|   | <b>Anacardiaceae</b>  | B      | 1   |
|   | <b>Anacardiaceae</b>  | L      | 1   |
|   | <b>Annonaceae</b>     | B      | 1   |
|   | <b>Asclepiadaceae</b> | O      | 2   |
|   | <b>Asteraceae</b>     | B      | 11  |
|   | <b>Asteraceae</b>     | O      | 3   |
|   | <b>Balanitaceae</b>   | B      | 1   |
|   | <b>Betulaceae</b>     | B      | 6   |
|   | <b>Bignoniaceae</b>   |        | 1   |
|   | <b>Bignoniaceae</b>   | B      | 4   |
|   | <b>Bixaceae</b>       |        | 1   |
|   | <b>Boraginaceae</b>   | B      | 1   |

Record: 1 of 83

# SQL query

```
sel: family-by-pltype : Select Q...  
SELECT *  
FROM gensp  
WHERE family = "Asteraceae";
```

| taxno  | genus         | species       | subsp | srcinf  | authname           | pltype | lngty | family     | syn    |
|--------|---------------|---------------|-------|---------|--------------------|--------|-------|------------|--------|
| ▶ 1400 | Eriocephalus  | ericoides     |       | Kew     | Druce              | B      | P     | Asteraceae | Tarch  |
| 1402   | Eriocephalus  | umbellatus    |       | Kew     | Cass.              | B      | P     | Asteraceae |        |
| 1461   | Felicia       | filifolia     |       | Kew v16 | (Vent.) Burtt-Davy | B      | P     | Asteraceae |        |
| 1462   | Felicia       | muricata      |       | Kew     | Nees               | B      | P     | Asteraceae |        |
| 1565   | Guizotia      | scabra        |       | Kew v16 | (Vis.) Chiov.      | O      | P     | Asteraceae |        |
| 1567   | Guizotia      | zavattarii    |       | Kew     | Lanza              | O      | P     | Asteraceae |        |
| 1607   | Hirpicium     | integrifolium |       | Kew     | Less.              | B      | P     | Asteraceae |        |
| 2847   | Osteospermum  | pachypteris   |       | Kew     | DC.                | B      | P     | Asteraceae | Tripte |
| 2848   | Osteospermum  | scariosum     |       | Kew     | DC.                | B      | P     | Asteraceae | Tripte |
| 2849   | Osteospermum  | sinuatum      |       | Kew v11 | DC. Norlindh.      | B      | P     | Asteraceae | Tripte |
| 2979   | Pentzia       | lanata        |       | Kew     | Hutchinson         | B      | P     | Asteraceae |        |
| 3680   | Tarchonanthus | camphoratus   |       | Kew     | Houtt.ex DC.       | B      | P     | Asteraceae | Brach  |
| 3980   | Vernonia      | amygdalina    |       | Kew     | Delile             | B      | P     | Asteraceae |        |
| 886    | Cichorium     | intybus       |       | Kew v1  | Linn.              | O      |       | Asteraceae |        |

Record: 1 of 14

# SQL query

```
freq: lat-lon : Select Query
SELECT iif(lat = "", "N", "Y") AS lat_present,
       iif(lon = "", "N", "Y") AS lon_present,
       count(*) AS cnt
FROM passport
GROUP BY iif(lat = "", "N", "Y"), iif(lon = "", "N", "Y");
```

|   | lat_present | lon_present | cnt         |
|---|-------------|-------------|-------------|
| ▶ | <b>N</b>    | <b>N</b>    | <b>6221</b> |
|   | <b>N</b>    | <b>Y</b>    | <b>3</b>    |
|   | <b>Y</b>    | <b>N</b>    | <b>244</b>  |
|   | <b>Y</b>    | <b>Y</b>    | <b>6735</b> |

Record: 1 of 4

# SQL query

```
freq: c-year : Select Query
SELECT year(cdate) AS year, count(*) AS cnt
FROM herb
GROUP BY year(cdate);
```

| year | cnt  |
|------|------|
| 1972 | 512  |
| 1973 | 481  |
| 1974 | 330  |
| 1975 | 504  |
| 1976 | 1278 |
| 1977 | 558  |
| 1978 | 476  |
| 1979 | 343  |
| 1980 | 411  |
| 1981 | 530  |
| 1982 | 188  |
| 1983 | 32   |
| 1984 | 31   |
| 1985 | 6    |
| 1986 | 3    |
| 1987 | 1    |
| 1988 | 3    |
| 1989 | 1    |
| 1995 | 2    |
| 1998 | 1    |
| 2962 | 1    |
| 4945 | 1    |

Record: 1

# SQL query

```
freq: weekday : Select Query
SELECT weekday(datetest) AS weekday, count(*) AS cnt
FROM germin
GROUP BY weekday(datetest);
```

|   | weekday | cnt  |
|---|---------|------|
| ▶ | 1       | 37   |
|   | 2       | 1667 |
|   | 3       | 1895 |
|   | 4       | 1196 |
|   | 5       | 2999 |
|   | 6       | 3337 |
|   | 7       | 147  |

Record: 1 of

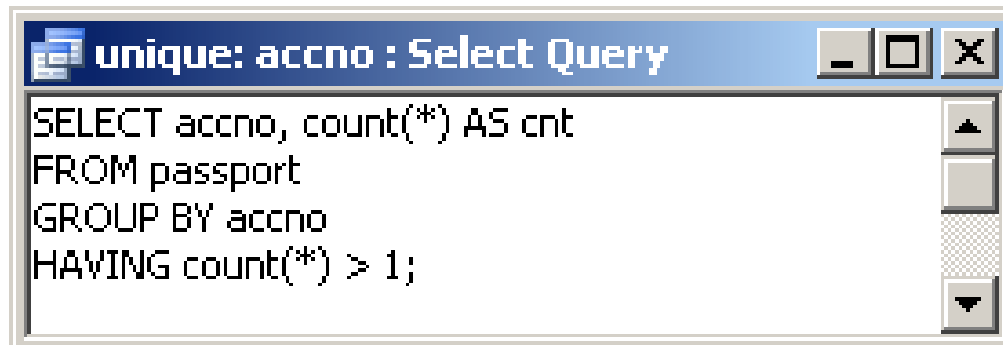
# SQL query

```
cnt: tests-per-date : Select Query
SELECT datetest, count(*) AS cnt
FROM germin
GROUP BY datetest
ORDER BY count(*);
```

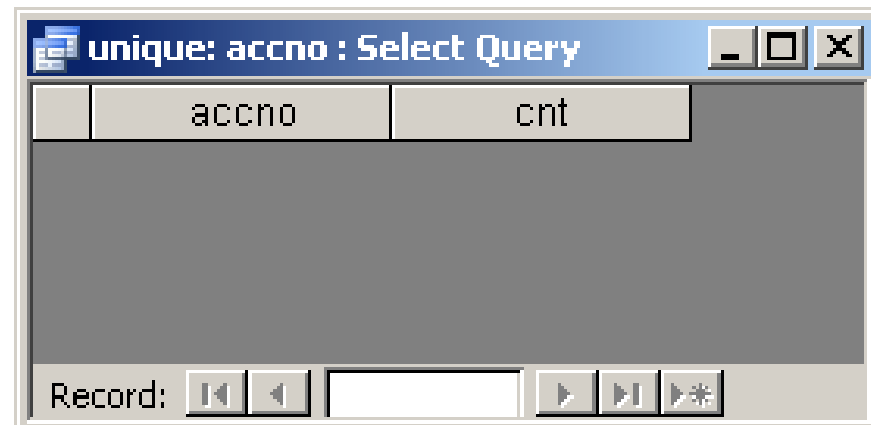
| datetest | cnt |
|----------|-----|
| 08/05/85 | 51  |
| 23/01/85 | 52  |
| 14/05/92 | 53  |
| 02/08/84 | 54  |
| 14/02/85 | 56  |
| 20/07/90 | 58  |
| 12/01/90 | 59  |
| 05/12/95 | 59  |
| 12/02/85 | 60  |
| 22/02/85 | 61  |
| 12/07/84 | 65  |
| 17/05/85 | 93  |

Record: 1

# SQL query

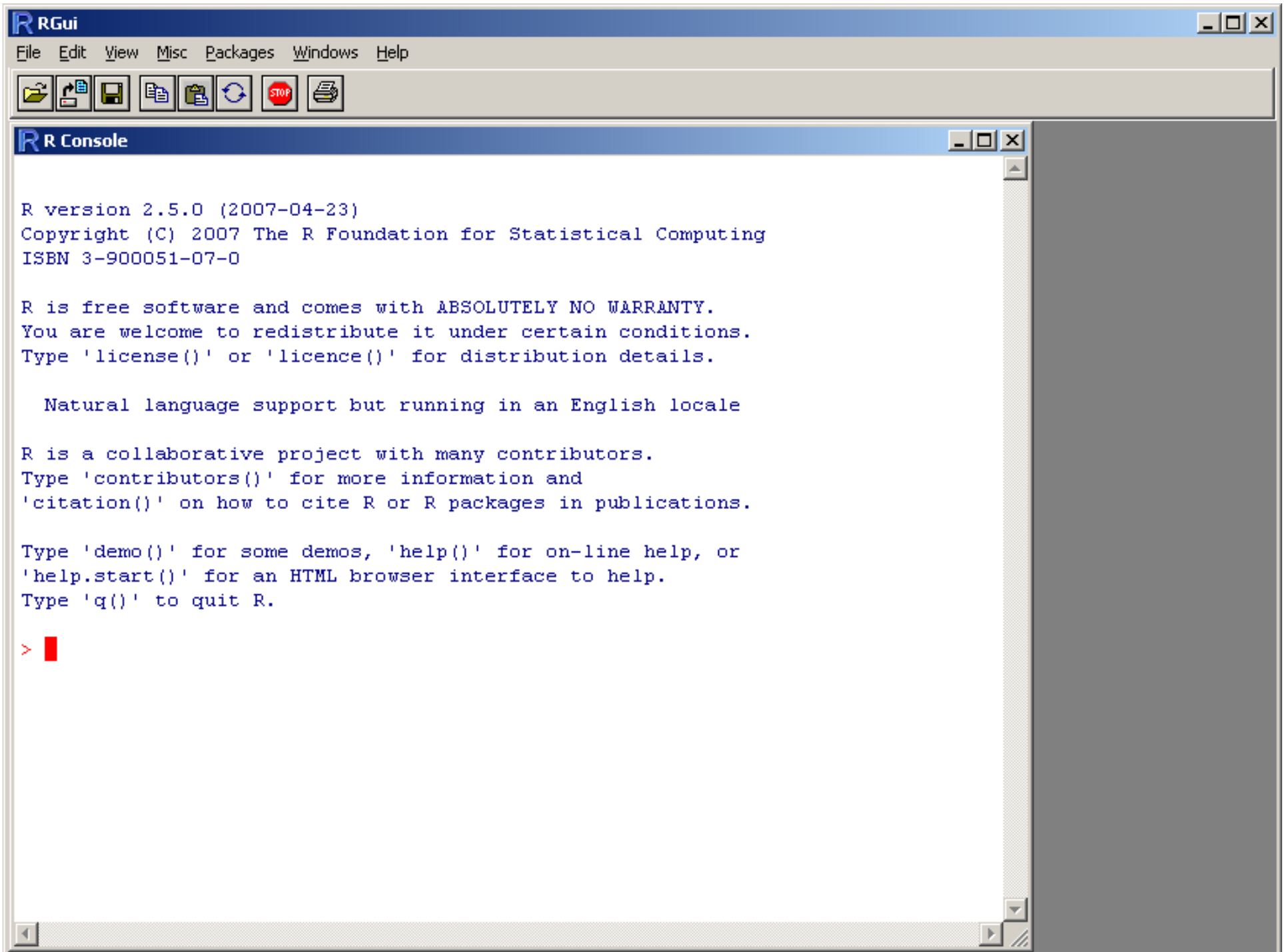


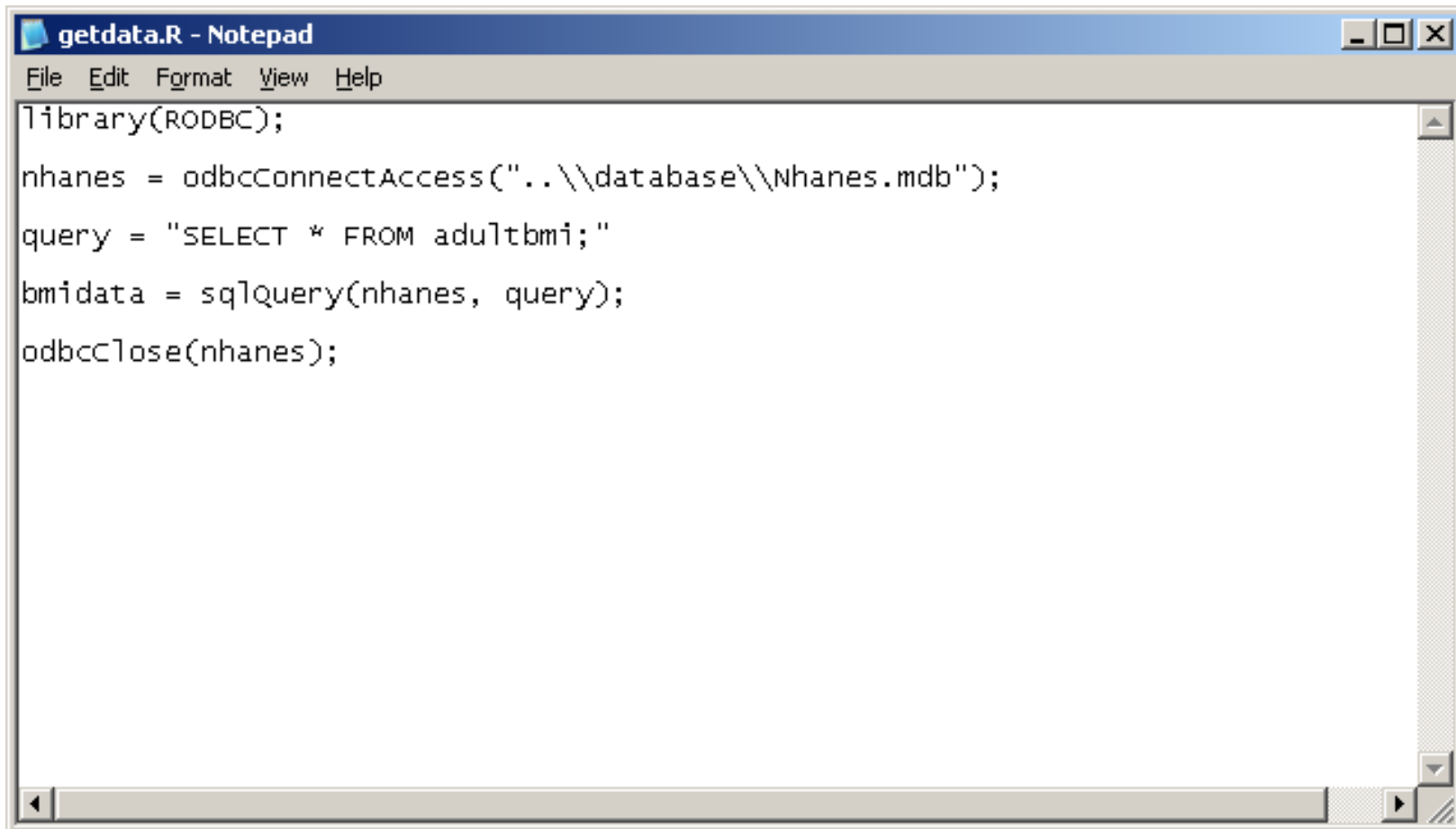
```
SELECT accno, count(*) AS cnt
FROM passport
GROUP BY accno
HAVING count(*) > 1;
```



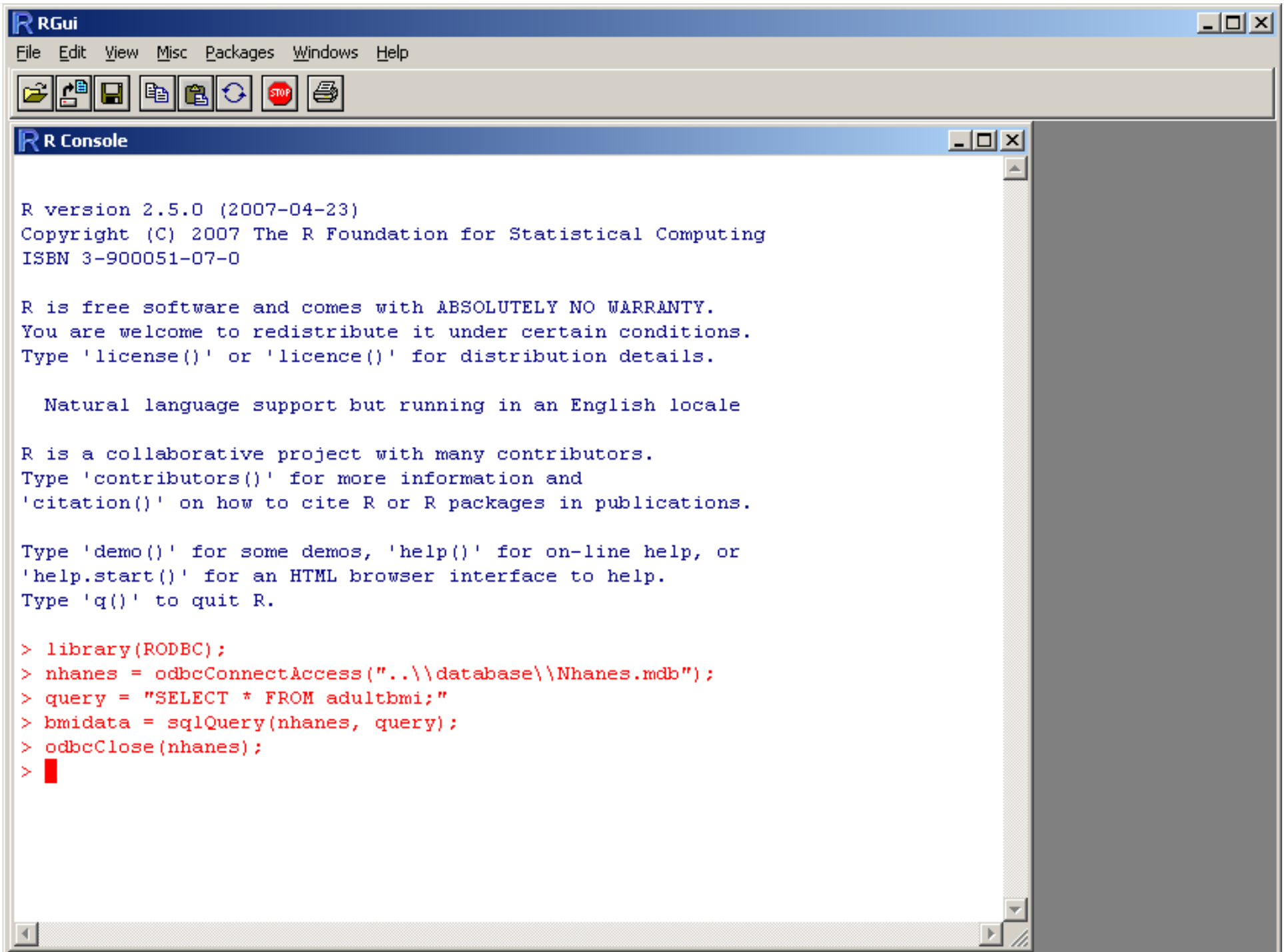
| accno | cnt |
|-------|-----|
|-------|-----|

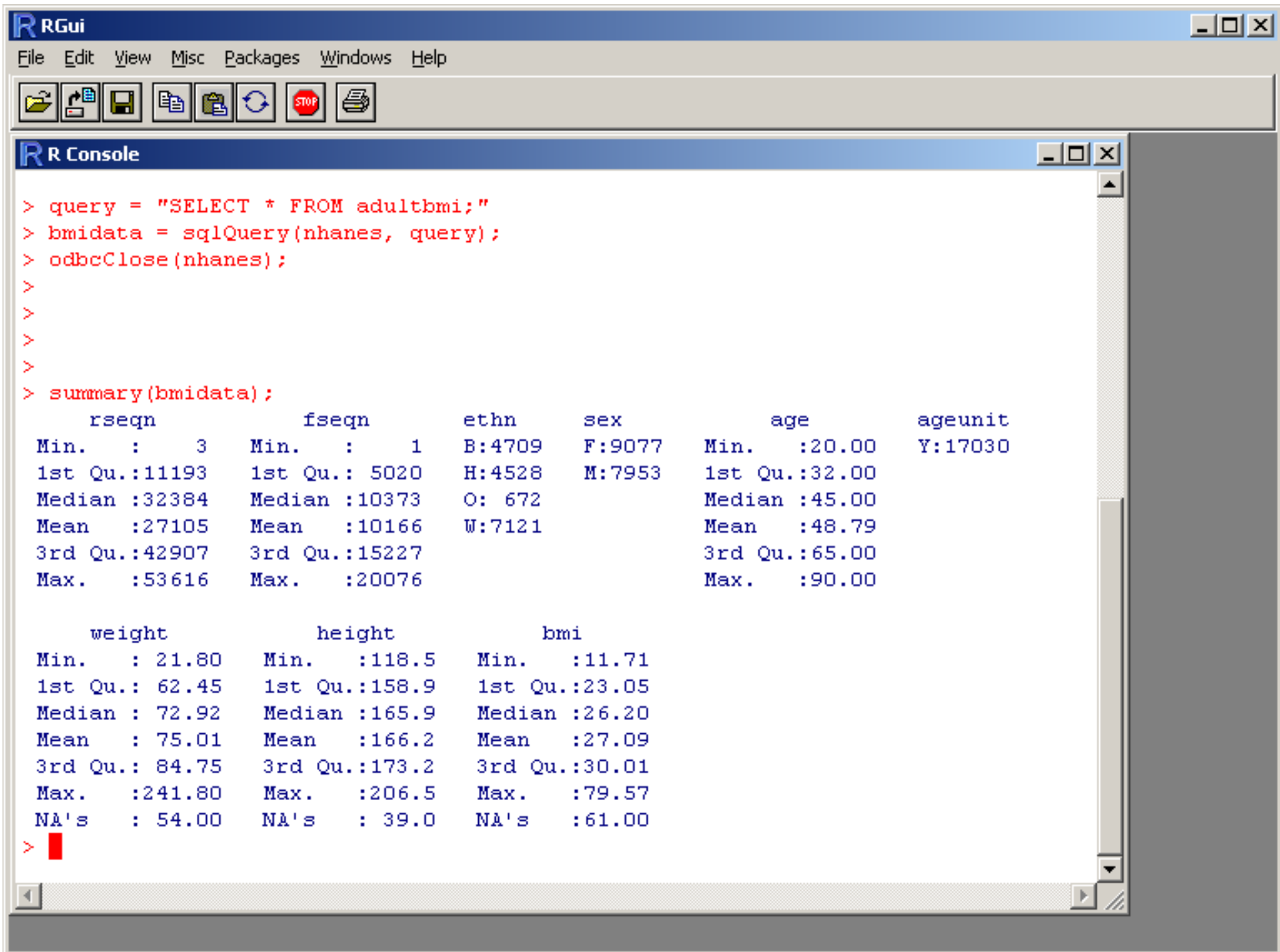
Record:  ⏪ ⏩ ⏴ ⏵ ⏶ ⏷





```
library(RODBC);  
nhanes = odbcConnectAccess("../database\\Nhanes.mdb");  
query = "SELECT * FROM adultbmi;"  
bmidata = sqlQuery(nhanes, query);  
odbcClose(nhanes);
```





RGui

File History Resize Windows

R Console

```
>
>
>
> summary(bmidata);
```

| rseqn         | fseqn         | ethn  |
|---------------|---------------|-------|
| Min. : 3      | Min. : 1      | B:470 |
| 1st Qu.:11193 | 1st Qu.: 5020 | H:452 |
| Median :32384 | Median :10373 | O: 67 |
| Mean :27105   | Mean :10166   | W:712 |
| 3rd Qu.:42907 | 3rd Qu.:15227 |       |
| Max. :53616   | Max. :20076   |       |

| weight         | height        |      |
|----------------|---------------|------|
| Min. : 21.80   | Min. :118.5   | Min. |
| 1st Qu.: 62.45 | 1st Qu.:158.9 | 1st  |
| Median : 72.92 | Median :165.9 | Medi |
| Mean : 75.01   | Mean :166.2   | Mean |
| 3rd Qu.: 84.75 | 3rd Qu.:173.2 | 3rd  |
| Max. :241.80   | Max. :206.5   | Max. |
| NA's : 54.00   | NA's : 39.0   | NA's |

```
>
>
>
> hist(bmidata$bmi);
> █
```

R Graphics: Device 2 (ACTIVE)

### Histogram of bmidata\$bmi

| bmi bin | Frequency |
|---------|-----------|
| 10-15   | 100       |
| 15-20   | 1200      |
| 20-25   | 5200      |
| 25-30   | 5500      |
| 30-35   | 2800      |
| 35-40   | 1000      |
| 40-45   | 400       |
| 45-50   | 200       |
| 50-55   | 100       |
| 55-60   | 50        |
| 60-65   | 20        |
| 65-70   | 10        |
| 70-75   | 5         |
| 75-80   | 2         |

RGui

File History Resize Windows

R Console

| rseqn         | fseqn         | ethn  |
|---------------|---------------|-------|
| Min. : 3      | Min. : 1      | B:470 |
| 1st Qu.:11193 | 1st Qu.: 5020 | H:452 |
| Median :32384 | Median :10373 | O: 67 |
| Mean :27105   | Mean :10166   | W:712 |
| 3rd Qu.:42907 | 3rd Qu.:15227 |       |
| Max. :53616   | Max. :20076   |       |

| weight         | height        | ethn |
|----------------|---------------|------|
| Min. : 21.80   | Min. :118.5   | Min. |
| 1st Qu.: 62.45 | 1st Qu.:158.9 | 1st  |
| Median : 72.92 | Median :165.9 | Medi |
| Mean : 75.01   | Mean :166.2   | Mean |
| 3rd Qu.: 84.75 | 3rd Qu.:173.2 | 3rd  |
| Max. :241.80   | Max. :206.5   | Max. |
| NA's : 54.00   | NA's : 39.0   | NA's |

```

>
>
>
> hist(bmidata$bmi);
>
>
>
> plot(bmidata$bmi ~ bmidata$age);
> █

```

R Graphics: Device 2 (ACTIVE)

bmidata\$bmi

bmidata\$age

RGui

File History Resize Windows

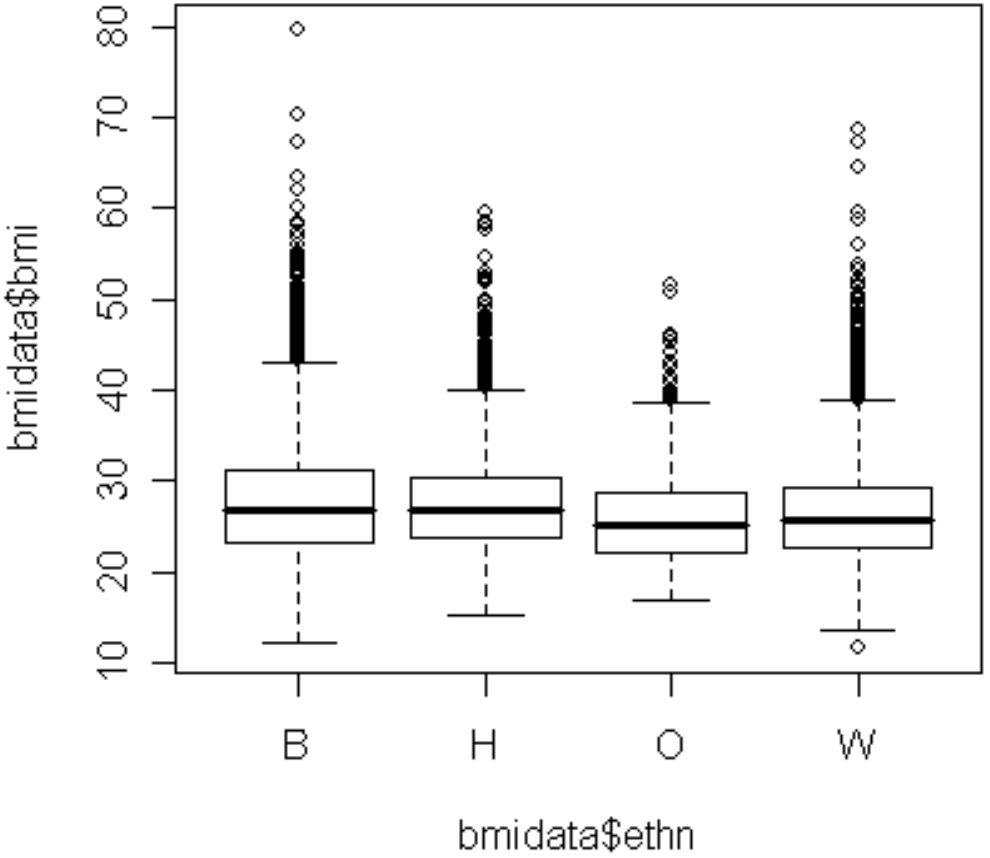
R Console

```
Mean :27105 Mean :10166 W:712
3rd Qu.:42907 3rd Qu.:15227
Max. :53616 Max. :20076
```

|          | weight  | height        |         |
|----------|---------|---------------|---------|
| Min.     | : 21.80 | Min. :118.5   | Min.    |
| 1st Qu.: | 62.45   | 1st Qu.:158.9 | 1st Qu. |
| Median : | 72.92   | Median :165.9 | Median  |
| Mean :   | 75.01   | Mean :166.2   | Mean    |
| 3rd Qu.: | 84.75   | 3rd Qu.:173.2 | 3rd Qu. |
| Max. :   | 241.80  | Max. :206.5   | Max.    |
| NA's :   | 54.00   | NA's : 39.0   | NA's    |

```
>
>
>
> hist(bmidata$bmi);
>
>
> plot(bmidata$bmi ~ bmidata$age);
>
>
>
> plot(bmidata$bmi ~ bmidata$ethn);
> █
```

R Graphics: Device 2 (ACTIVE)



A box plot showing the distribution of BMI (bmidata\$bmi) across four ethnic groups (B, H, O, W). The y-axis ranges from 10 to 80. The x-axis is labeled 'bmidata\$ethn'. The plot shows that BMI is generally higher for groups B and W compared to H and O. Group B has the highest median BMI, followed by group W, then group H, and group O has the lowest median BMI. All groups show a similar spread of data, with many outliers above the upper whisker.

| ethn | Min | Q1 | Median | Q3 | Max |
|------|-----|----|--------|----|-----|
| B    | 12  | 23 | 27     | 31 | 43  |
| H    | 15  | 24 | 27     | 30 | 40  |
| O    | 17  | 22 | 25     | 29 | 39  |
| W    | 13  | 23 | 26     | 29 | 39  |



RGui

File History Resize Windows

- Save as ▶
  - Metafile...
  - Postscript...
  - PDF...
  - Png...
  - Bmp...
  - Jpeg ▶
    - 50% quality...
    - 75% quality...
    - 100% quality...
- Copy to the clipboard ▶
- Print... CTRL+P
- close Device

R Graphics: Device 2 (ACTIVE)

|                  | age         | age         | age         | age         |
|------------------|-------------|-------------|-------------|-------------|
|                  | B           | H           | O           | W           |
| Percent of Total | [Histogram] | [Histogram] | [Histogram] | [Histogram] |
|                  | bmi         | bmi         | bmi         | bmi         |
|                  | B           | H           | O           | W           |
|                  | [Histogram] | [Histogram] | [Histogram] | [Histogram] |

bmi + age

```
> library(lattice);  
> histogram(~ bmi + age | ethn, data=bmidata, nint=30, type="percent");  
> █
```

RGui

File Edit View Misc Packages Windows Help

R Console

```
>
>
>
>
>
>
> bmidata$ethn <- factor(bmidata$ethn);
> bmidata$sex <- factor(bmidata$sex);
>
> lmbmi <- lm(formula = bmi ~ ethn + sex + ethn:sex + age, data = bmidata);
>
>
>
> anova(lmbmi);
Analysis of Variance Table

Response: bmi

```

|           | Df    | Sum Sq | Mean Sq | F value | Pr(>F)    |     |
|-----------|-------|--------|---------|---------|-----------|-----|
| ethn      | 3     | 7745   | 2582    | 78.044  | < 2.2e-16 | *** |
| sex       | 1     | 3536   | 3536    | 106.879 | < 2.2e-16 | *** |
| age       | 1     | 1806   | 1806    | 54.594  | 1.551e-13 | *** |
| ethn:sex  | 3     | 5049   | 1683    | 50.882  | < 2.2e-16 | *** |
| Residuals | 16960 | 561033 | 33      |         |           |     |

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
>
>
>
> █
```

RGui

File Edit View Misc Packages Windows Help

R Console

```
> summary(lmbmi);
```

Call:  
lm(formula = bmi ~ ethn + sex + ethn:sex + age, data = bmidata)

Residuals:

| Min      | 1Q      | Median  | 3Q     | Max     |
|----------|---------|---------|--------|---------|
| -17.1063 | -3.9092 | -0.7946 | 2.8905 | 50.7651 |

Coefficients:

|             | Estimate  | Std. Error | t value | Pr(> t ) |     |
|-------------|-----------|------------|---------|----------|-----|
| (Intercept) | 28.210748 | 0.153231   | 184.106 | < 2e-16  | *** |
| ethnH       | -0.897186 | 0.165957   | -5.406  | 6.53e-08 | *** |
| ethnO       | -2.337631 | 0.315364   | -7.412  | 1.30e-13 | *** |
| ethnW       | -2.893266 | 0.149123   | -19.402 | < 2e-16  | *** |
| sexM        | -2.477917 | 0.168792   | -14.680 | < 2e-16  | *** |
| age         | 0.017475  | 0.002359   | 7.407   | 1.36e-13 | *** |
| ethnH:sexM  | 1.406552  | 0.240617   | 5.846   | 5.14e-09 | *** |
| ethnO:sexM  | 1.384642  | 0.480400   | 2.882   | 0.00395  | **  |
| ethnW:sexM  | 2.673760  | 0.217262   | 12.307  | < 2e-16  | *** |

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.751 on 16960 degrees of freedom  
(61 observations deleted due to missingness)  
Multiple R-Squared: 0.03131, Adjusted R-squared: 0.03086  
F-statistic: 68.53 on 8 and 16960 DF, p-value: < 2.2e-16

```
> █
```