Project 3: The first part

Cassie Boylan, DH Kim, Alexis Mekueko

2020-10-10

Data Source and Loading

Packages used:

```
library(readxl)
library(httr)
```

E-Commerce Data:

```
retailURL <-
    "http://archive.ics.uci.edu//ml//machine-learning-databases//00502//online_retail_II.xlsx"
```

Loading and Combining Data

```
GET(retailURL, write_disk(tempFileName <- tempfile(fileext = ".xlsx")))
```

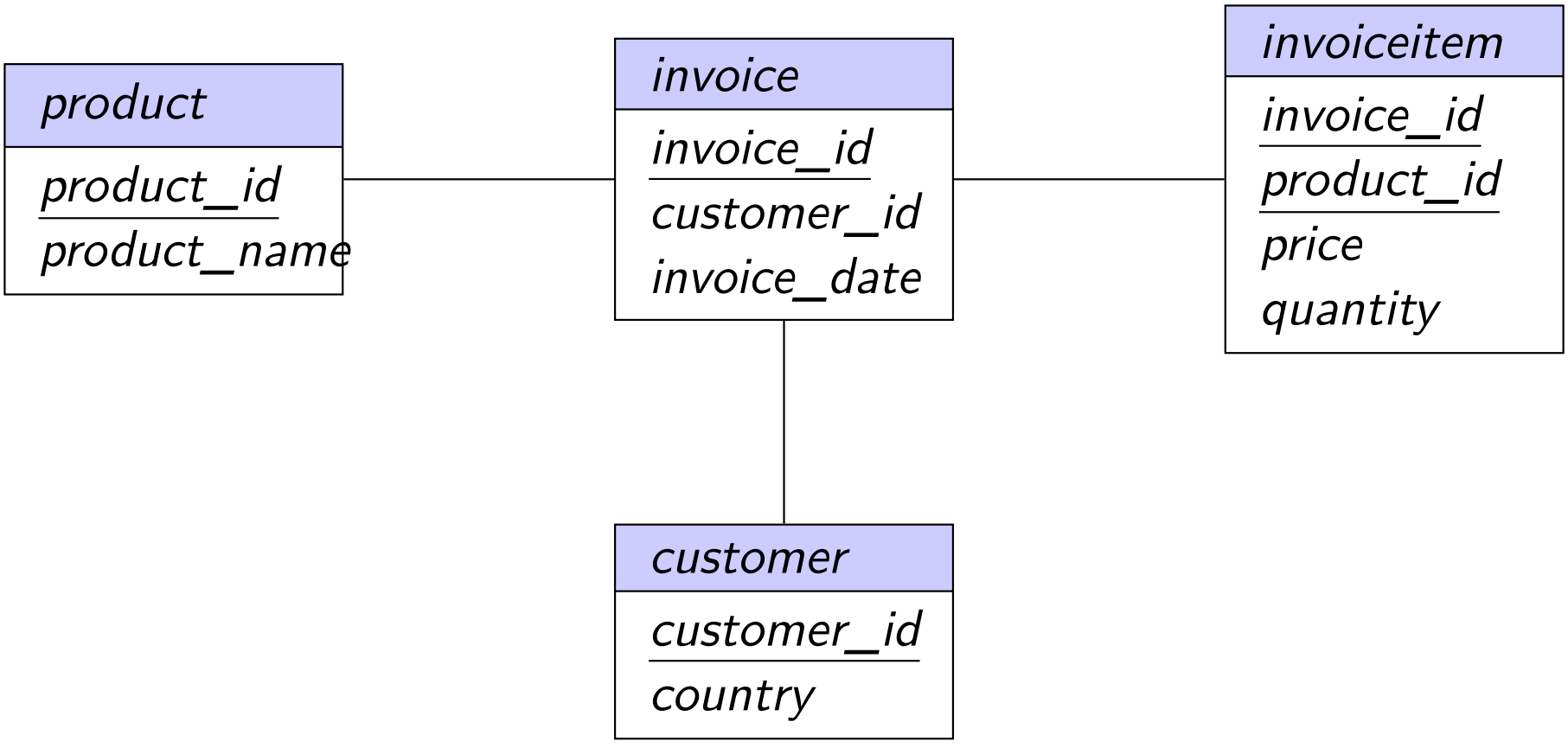
```
retail_sheet_2009 <- read_excel(tempFileName, sheet = "Year 2009-2010")
retail_sheet_2010 <- read_excel(tempFileName, sheet = "Year 2010-2011")
retaildf <- rbind(retail_sheet_2009, retail_sheet_2010)
```

Description of Data

```
retaildf
```

```
## # A tibble: 1,067,371 x 8
##    Invoice StockCode Description Quantity InvoiceDate         Price
##    <chr>   <chr>     <chr>          <dbl> <dttm>              <dbl>
##  1 489434  85048     "15CM CHRI~       12 2009-12-01 07:45:00  6.95
##  2 489434  79323P    "PINK CHER~       12 2009-12-01 07:45:00  6.75
##  3 489434  79323W    "WHITE CHE~       12 2009-12-01 07:45:00  6.75
##  4 489434  22041     "RECORD FR~       48 2009-12-01 07:45:00  2.1
##  5 489434  21232     "STRAWBERR~       24 2009-12-01 07:45:00  1.25
##  6 489434  22064     "PINK DOUG~       24 2009-12-01 07:45:00  1.65
##  7 489434  21871     "SAVE THE ~       24 2009-12-01 07:45:00  1.25
##  8 489434  21523     "FANCY FON~       10 2009-12-01 07:45:00  5.95
##  9 489435  22350     "CAT BOWL"        12 2009-12-01 07:46:00  2.55
## 10 489435  22349     "DOG BOWL ~       12 2009-12-01 07:46:00  3.75
## # ... with 1,067,361 more rows, and 2 more variables: `Customer ID` <dbl>,
## #   Country <chr>
```

E-R Diagram


product
product_id
product_name
invoice
invoice_id
customer_id
invoice_date
invoiceitem
invoice_id
product_id
price
quantity
customer
customer_id
country