

Homework 1

Andey Nunes, MS
additional team member name(s) here
January 10, 2019

Document Setup

The first step for this week is to set up the R Markdown document options.

Next step, load the data sets for the homework and summarize.

```
catalog <- read_excel("catalog.xls")
customers <- read_excel("customers.xls")
order_lines <- read_excel("order_lines.xls")
```

```
## New names:
## * `` -> ``.2`
```

```
orders <- read_excel("orders.xls")
```

Summary tables

```
catalog_summary <- summary(catalog)

pander(catalog_summary, caption = "catalog summary table")
```

Table 1: catalog summary table (continued below)

id	product_code	catalog_price	category1
Min. : 307	Length:761	Min. : 0	Length:761
1st Qu.: 525	Class :character	1st Qu.: 18	Class :character
Median : 728	Mode :character	Median : 34	Mode :character
Mean : 725	NA	Mean : 49	NA
3rd Qu.: 930	NA	3rd Qu.: 57	NA
Max. :1125	NA	Max. :654	NA

manufact_id	vendor_id	name
Min. :0.0	Min. :0.0	Length:761
1st Qu.:1.0	1st Qu.:1.0	Class :character
Median :1.0	Median :1.0	Mode :character
Mean :1.2	Mean :1.2	NA
3rd Qu.:1.0	3rd Qu.:1.0	NA
Max. :8.0	Max. :8.0	NA

```
customers_summary <- summary(customers)

pander(customers_summary, caption = "customers summary table")
```

Table 3: customers summary table (continued below)

cust_id	merchant_id	firstName	lastName
Min. :10000	Min. :1.00	Length:22070	Length:22070
1st Qu.:15930	1st Qu.:1.00	Class :character	Class :character
Median :21448	Median :1.00	Mode :character	Mode :character
Mean :21408	Mean :1.05	NA	NA
3rd Qu.:26965	3rd Qu.:1.00	NA	NA
Max. :32482	Max. :2.00	NA	NA

Table 4: Table continues below

bt_city	bt_state	bt_country	bt_zip
Length:22070	Length:22070	Length:22070	Length:22070
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
NA	NA	NA	NA
NA	NA	NA	NA
NA	NA	NA	NA

cc_type	custcode
Length:22070	Length:22070
Class :character	Class :character
Mode :character	Mode :character
NA	NA
NA	NA
NA	NA

```
order_lines_summary <- summary(order_lines)
pander(order_lines_summary, caption = "order_lines summary table")
```

Table 6: order_lines summary table

Sum of Shipped Total	..2
Length:1356	Length:1356
Class :character	Class :character
Mode :character	Mode :character

```
orders_summary <- summary(orders)
pander(orders_summary, caption = "orders summary table")
```

Table 7: orders summary table (continued below)

order_id	merchant_id	order_date	po_number
Min. :14000	Min. :1.00	Min. :2003-10-10 00:00:00	Length:23256
1st Qu.:20134	1st Qu.:1.00	1st Qu.:2006-04-28 00:00:00	Class :character

order_id	merchant_id	order_date	po_number
Median :25948	Median :1.00	Median :2007-07-02 00:00:00	Mode :character
Mean :25918	Mean :1.05	Mean :2007-08-11 16:51:42	NA
3rd Qu.:31761	3rd Qu.:1.00	3rd Qu.:2008-12-19 00:00:00	NA
Max. :37575	Max. :2.00	Max. :2011-01-21 00:00:00	NA

Table 8: Table continues below

cust_id	order_status	ship_method	items_amount
Min. : 0	Length:23256	Length:23256	Min. : 0
1st Qu.:15778	Class :character	Class :character	1st Qu.: 28
Median :21302	Mode :character	Mode :character	Median : 48
Mean :21295	NA	NA	Mean : 73
3rd Qu.:26849	NA	NA	3rd Qu.: 80
Max. :32482	NA	NA	Max. :9590

Table 9: Table continues below

amt_bracket	total_weight	total_ship	total_hand	total_tax
Length:23256	Min. : 0	Min. : 0	Min. :0	Min. :0
Class :character	1st Qu.: 1	1st Qu.: 7	1st Qu.:0	1st Qu.:0
Mode :character	Median : 2	Median : 8	Median :0	Median :0
NA	Mean : 3	Mean : 11	Mean :0	Mean :0
NA	3rd Qu.: 3	3rd Qu.: 10	3rd Qu.:0	3rd Qu.:0
NA	Max. :483	Max. :631	Max. :0	Max. :0

Table 10: Table continues below

total_amount	order_status_date	send_inv_to_bill	coupon_code
Min. : 6	Min. :2003-10-10 00:00:00	Min. :0.00	Mode:logical
1st Qu.: 36	1st Qu.:2006-05-30 18:00:00	1st Qu.:0.00	NA's:23256
Median : 57	Median :2007-07-12 00:00:00	Median :0.00	NA
Mean : 84	Mean :2007-08-21 21:51:27	Mean :0.05	NA
3rd Qu.: 94	3rd Qu.:2008-12-26 00:00:00	3rd Qu.:0.00	NA
Max. :9590	Max. :2011-01-21 00:00:00	Max. :1.00	NA

spec_instr
Mode:logical
NA's:23256
NA
NA
NA
NA

column names (variables) | assign a type: “question”, “answer”, or “link” | variable class | count missing values | range = max - min |

This section is for building some custom functions that will come in handy later

```
countNA <- function(x) { sum(is.na(x)) }  
get_range <- function(x) {diff(range(x))}
```

Homework Questions

References