

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT: Data Analysis with R

Aim: Sorting data using arrange() in R

```
> # Load your dataset
> data <- read_csv("flavors_of_cacao.csv", show_col_types = FALSE)

> # Convert Cocoa Percent into numeric (e.g., "70%" -> 70)
> data$cocoaPercentNum <- as.numeric(gsub("%", "", data$cocoaPercent))
> # Quick view
> head(data)
# A tibble: 6 x 10
  Company \n(Maker-if ... Specific Bean origin_ REF Review\ndate `Cocoa\nPercent` `Company\nLocation` Rating
  <chr> <chr> <dbl> <dbl> <chr> <chr> <dbl>
1 A. Morin Agua Grande 1876 2016 63% France 3.75
2 A. Morin Kpime 1676 2015 70% France 2.75
3 A. Morin Atsane 1676 2015 70% France 3
4 A. Morin Akata 1680 2015 70% France 3.5
5 A. Morin Quilla 1704 2015 70% France 3.5
6 A. Morin Carenero 1315 2014 70% France 2.75
# i abbreviated names: 'Company \n(Maker-if known)', 'Specific Bean origin\nor Bar Name'
# i 3 more variables: 'Bean\ntype' <chr>, 'Broad Bean\norigin' <chr>, 'CocoaPercentNum' <dbl>
> sorted_rating <- data %>%
+ arrange(desc(Rating))
> head(sorted_rating)
# A tibble: 5 x 10
  Company \n(Maker-if ... Specific Bean origin_ REF Review\ndate `Cocoa\nPercent` `Company\nLocation` Rating
  <chr> <chr> <dbl> <dbl> <chr> <chr> <dbl>
1 Callebaut Baking 141 2007 70% Belgium 1
2 Claudio corallo Principe 252 2008 100% Sao Tome 1
3 Cote d' or (kraft) Sensations Intense 48 2006 70% Belgium 1
4 Neuhaus (Callebaut) Dark 135 2007 73% Belgium 1
5 Bonnat One Hundred 81 2006 100% France 1.5
# i abbreviated names: 'Company \n(Maker-if known)', 'Specific Bean origin\nor Bar Name'
# i 3 more variables: 'Bean\ntype' <chr>, 'Broad Bean\norigin' <chr>, 'CocoaPercentNum' <dbl>
> sorted_cocoa_desc <- data %>%
+ arrange(desc(cocoaPercentNum))
> head(sorted_cocoa_desc, 5)
# A tibble: 5 x 10
```

Company (Maker-if known)	Specific Bean Origin or Bar Name	REF	Review Date	Cocoa Percent	Company Location	Rating	Bean Type	Broad Bean Origin	CocoaPercentNum
1 A. Morin	Agua Grande	1876	2016	63%	France	3.75	NA	Sao Tome	63
2 A. Morin	Kpime	1676	2015	70%	France	2.75		Togo	70
3 A. Morin	Atsane	1676	2015	70%	France	3.00	NA	Togo	70
4 A. Morin	Akata	1680	2015	70%	France	3.50		Togo	70
5 A. Morin	Quilla	1704	2015	70%	France	3.50	NA	Peru	70
6 A. Morin	Carenero	1315	2014	70%	France	2.75	Criollo	Venezuela	70
7 A. Morin	Cuba	1315	2014	70%	France	3.50		Cuba	70
8 A. Morin	Sur del Lago	1315	2014	70%	France	3.50	Criollo	Venezuela	70
9 A. Morin	Puerto Cabello	1319	2014	70%	France	3.75	Criollo	Venezuela	70
10 A. Morin	Pabino	1319	2014	70%	France	4.00		Peru	70
11 A. Morin	Panama	1011	2013	70%	France	2.75		Panama	70
12 A. Morin	Madagascar	1011	2013	70%	France	3.00	Criollo	Madagascar	70
13 A. Morin	Brazil	1011	2013	70%	France	3.25		Brazil	70
14 A. Morin	Ecatepec	1011	2013	70%	France	3.75		Ecuador	70
15 A. Morin	Colombia	1015	2013	70%	France	2.75		Colombia	70

```
Showing 1 to 15 of 1,795 entries, 10 total columns

Console Terminal Background Jobs
R 4.5.2 ~ /
> Salgado Rio Arriba 292 2008 70% Argentina
> Compania de chocolate (Salgado) Sur del Lago 579 2010 70% Argentina
# i abbreviated names: 'Company \n(Maker-if known)', 'Specific Bean origin\nor Bar Name'
# i 4 more variables: Rating <dbl>, 'Bean\ntype' <chr>, 'Broad Bean\norigin' <chr>, 'CocoaPercentNum' <dbl>
> high_cocoa_sorted <- data %>%
+ filter(CocoaPercentNum > 80) %>%
+ arrange(Rating)
> head(high_cocoa_sorted, 5)
```

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT: Data Analysis with R

Company (Maker-if known)	Specific Bean Origin or Bar Name	REF	Review Date	Cocoa Percent	Company Location	Rating	Bean Type	Broad Bean Origin	CocoaPercentNum
1 Claudio Corallo	Principe	252	2008	100%	Sao Tome	1.00	Forastero	Sao Tome & Principe	100
2 Bonnat	One Hundred	81	2006	100%	France	1.50			100
3 Bouga Cacao (Tullcorp)	El Oro, Hacienda de Oro	341	2009	100%	Ecuador	1.50	Forastero (Arriba)	Ecuador	100
4 Cacaoyere (Ecuadoriana)	Pichincha	259	2008	91%	Ecuador	1.50	Forastero (Arriba)	Ecuador	91
5 Coppeneur	Ecuador Puristique	817	2012	100%	Germany	1.50		Ecuador	100
6 S.A.I.D.	100 percent	615	2010	100%	Italy	1.50			100
7 Vairhona	Le Noir Extra Amer	147	2007	85%	France	1.50			85
8 Artisan du Chocolat	Venezuela	466	2010	100%	U.K.	1.75	N/A	Venezuela	100
9 Hotel Chocolat	Island Growers, 2012, 120hr c.	1113	2013	100%	U.K.	1.75	Trinitario	St. Lucia	100
10 French Broad	Palo Blanco w/ panela, Chulucanas	887	2012	81%	U.S.A.	2.00		Peru	81
11 Habitual	one hundred	1205	2014	100%	Canada	2.00			100
12 Michel Cluizel	Carre Grand Noir	24	2006	85%	France	2.00			85
13 Michel Cluizel	Noir Infini	81	2006	99%	France	2.00			99
14 Pacari	Raw	266	2008	100%	Ecuador	2.00	Forastero (Arriba)	Ecuador	100
15 Scharffen Berger	Extra Dark	15	2006	82%	U.S.A.	2.00			82

Showing 1 to 15 of 111 entries, 10 total columns

```

R - R 4.5.2 - ~/
1 100%      1 Sao Tome
2 100%      1.5 France
3 100%      1.5 Ecuador
4 91%       1.5 Ecuador
5 100%      1.5 Germany
> view(data)
> view(high_cocoa_sorted)

```

Company (Maker-if known)	Specific Bean Origin or Bar Name	REF	Review Date	Cocoa Percent	Company Location	Rating	Bean Type	Broad Bean Origin	CocoaPercentNum
1 Chocolate Makers	Peru, Awagum bar	1530	2015	80%	Amsterdam	3.75	Criollo	Peru	80
2 Chocolate Makers	Congo, Gorilla bar	1530	2015	68%	Amsterdam	3.50	Forastero	Congo	68
3 Metropolitan	South America	1688	2015	70%	Amsterdam	3.50	Trinitario	South America	70
4 Chocolate Makers	Tres Hombres	1530	2015	75%	Amsterdam	3.25	Trinitario	Dominican Republic	75
5 Compania de Chocolate (Saigado)	Ocumare	292	2008	70%	Argentina	3.75	Criollo	Venezuela	70
6 Saigado	Bahia Superior	288	2008	70%	Argentina	3.50	Forastero	Brazil	70
7 Saigado	Esmeraldas	288	2008	70%	Argentina	3.50	Forastero (Nacional)	Ecuador	70
8 Saigado	Carenero Superior	288	2008	70%	Argentina	3.50	Trinitario	Venezuela	70
9 Saigado	Rio Arriba	292	2008	70%	Argentina	3.50	Forastero (Arriba) ASSS	Ecuador	70
10 Compania de Chocolate (Saigado)	Sur del Lago	579	2010	70%	Argentina	3.25		Venezuela	70
11 Compania de Chocolate (Saigado)	Moxos	292	2008	72%	Argentina	3.25		Bolivia	72
12 Compania de Chocolate (Saigado)	Carenero Superior	292	2008	80%	Argentina	2.75	Trinitario	Venezuela	80
13 Compania de Chocolate (Saigado)	Esmeraldas	296	2008	88%	Argentina	2.75	Forastero (Nacional)	Ecuador	88
14 Matale	Somia, 2013	1177	2013	68%	Australia	4.00	Trinitario	Madagascar	68
15 Matale	Kulili P, 2013	1177	2013	74%	Australia	4.00	Trinitario	Papua New Guinea	74

Showing 1 to 15 of 1,795 entries, 10 total columns

```

R - R 4.5.2 - ~/
2 100%      1.5 France
3 100%      1.5 Ecuador
4 91%       1.5 Ecuador
5 100%      1.5 Germany
> view(data)
> view(high_cocoa_sorted)
> view(multi_sorted)

```

# SHETH L.U.J. AND SIR M.V. COLLEGE

## SUBJECT: Data Analysis with R

The screenshot displays the RStudio interface with the following components:

- Top Panel (Data Table):** Shows a table with 10 columns and 15 rows of data. The columns are: Company (Maker-if known), Specific Bean Origin or Bar Name, REF, Review Date, Cocoa Percent, Company Location, Rating, Bean Type, Broad Bean Origin, and CocoaPercentNum. The rows list various chocolate makers and their products, such as Artisan du Chocolat, Bonnat, Bouga Cacao, etc.
- Bottom Panel (R Console):** Contains R code for data manipulation. The code includes:
  - Sorting the data by cocoa percent in descending order: `sorted_cocoa_desc <- data %>% arrange(desc(CocoaPercentNum))`
  - Viewing the first 5 rows: `head(sorted_cocoa_desc, 5)`
  - Sorting the data by rating in descending order: `multi_sorted <- data %>% arrange(desc(Rating))`
  - Viewing the first 10 rows: `head(multi_sorted, 10)`
  - Filtering the data for high cocoa content (CocoaPercentNum > 80): `high_cocoa_sorted <- data %>% filter(CocoaPercentNum > 80)`
  - Viewing the first 5 rows of the filtered data: `head(high_cocoa_sorted, 5)`
- Right Panel (Files):** Shows a list of files in the current directory, including `Cleaned_data.csv`, `flavors_of_cacao.csv`, and `GROW S103 FINTECH.pptx`.