Chipotle Homework

1. Look at the head and the tail of chipotle.tsv in the data subdirectory of this repo. Think for a minute about how the data is structured. What do you think each column means? What do you think each row means? Tell me! (If you're unsure, look at more of the file contents.)

danielles-air:data Danielle$ head chipotle.tsv

order\_id quantity item\_name choice\_description item\_price

1 1 Chips and Fresh Tomato Salsa NULL $2.39

1 1 Izze [Clementine] $3.39

1 1 Nantucket Nectar [Apple] $3.39

1 1 Chips and Tomatillo-Green Chili Salsa NULL $2.39

2 2 Chicken Bowl [Tomatillo-Red Chili Salsa (Hot), [Black Beans, Rice, Cheese, Sour Cream]] $16.98

3 1 Chicken Bowl [Fresh Tomato Salsa (Mild), [Rice, Cheese, Sour Cream, Guacamole, Lettuce]] $10.98

3 1 Side of Chips NULL $1.69

4 1 Steak Burrito [Tomatillo Red Chili Salsa, [Fajita Vegetables, Black Beans, Pinto Beans, Cheese, Sour Cream, Guacamole, Lettuce]] $11.75

4 1 Steak Soft Tacos [Tomatillo Green Chili Salsa, [Pinto Beans, Cheese, Sour Cream, Lettuce]] $9.25

danielles-air:data Danielle$ tail chipotle.tsv

1831 1 Carnitas Bowl [Fresh Tomato Salsa, [Fajita Vegetables, Rice, Black Beans, Cheese, Sour Cream, Lettuce]] $9.25

1831 1 Chips NULL $2.15

1831 1 Bottled Water NULL $1.50

1832 1 Chicken Soft Tacos [Fresh Tomato Salsa, [Rice, Cheese, Sour Cream]] $8.75

1832 1 Chips and Guacamole NULL $4.45

1833 1 Steak Burrito [Fresh Tomato Salsa, [Rice, Black Beans, Sour Cream, Cheese, Lettuce, Guacamole]] $11.75

1833 1 Steak Burrito [Fresh Tomato Salsa, [Rice, Sour Cream, Cheese, Lettuce, Guacamole]] $11.75

1834 1 Chicken Salad Bowl [Fresh Tomato Salsa, [Fajita Vegetables, Pinto Beans, Guacamole, Lettuce]] $11.25

1834 1 Chicken Salad Bowl [Fresh Tomato Salsa, [Fajita Vegetables, Lettuce]] $8.75

1834 1 Chicken Salad Bowl [Fresh Tomato Salsa, [Fajita Vegetables, Pinto Beans, Lettuce]] $8.75

danielles-air:data Danielle$

The columns show the order ID, quantity,item name, description, and price

The rows seem to go in chronological order.

1. How many orders do there appear to be?

Looking at the tail, the last item is numbered as order: 1834

1. How many lines are in this file?

Using the WC command:

danielles-air:data Danielle$ wc -l chipotle.tsv

4623 chipotle.tsv

I’m seeing 4623 lines in the file.

1. Which burrito is more popular, steak or chicken?

Steak: 368

Chicken: 553

danielles-air:data Danielle$ cat chipotle.tsv | grep "Chicken Burrito" chipotle.tsv | wc -l

553

danielles-air:data Danielle$ cat chipotle.tsv | grep "Steak Burrito" chipotle.tsv | wc -l

368

danielles-air:data Danielle$

(Chicken is more popular)

danielles-air:data Danielle$ grep "Steak Burrito" chipotle.tsv

danielles-air:data Danielle$ grep "Chicken Burrito" chipotle.tsv

danielles-air:data Danielle$ grep "Chicken Burrito" chipotle.tsv | tail

1799 1 Chicken Burrito [Roasted Chili Corn Salsa, [Rice, Black Beans, Cheese]] $8.75

1800 1 Chicken Burrito [Fresh Tomato Salsa, [Rice, Black Beans, Sour Cream, Cheese, Lettuce]] $8.75

1801 1 Chicken Burrito [Roasted Chili Corn Salsa, [Fajita Vegetables, Rice, Cheese, Lettuce]] $8.75

1802 1 Chicken Burrito [Tomatillo Red Chili Salsa, [Fajita Vegetables, Rice, Black Beans, Cheese, Sour Cream, Guacamole, Lettuce]] $11.25

1806 1 Chicken Burrito [Fresh Tomato Salsa, [Fajita Vegetables, Rice]]$8.75

1811 1 Chicken Burrito [Tomatillo Green Chili Salsa, [Fajita Vegetables, Rice, Black Beans, Cheese, Sour Cream, Guacamole, Lettuce]] $11.25

1812 1 Chicken Burrito [Fresh Tomato Salsa, [Rice, Black Beans, Cheese, Sour Cream, Lettuce]] $8.75

1812 1 Chicken Burrito [Tomatillo Red Chili Salsa, [Rice, Cheese, Guacamole, Lettuce]] $11.25

1816 1 Chicken Burrito [Roasted Chili Corn Salsa, [Pinto Beans, Cheese, Sour Cream]] $8.75

1817 1 Chicken Burrito [Fresh Tomato Salsa, [Fajita Vegetables, Rice]]$8.75

danielles-air:data Danielle$

1. Do chicken burritos more often have black beans or pinto beans?

danielles-air:data Danielle$ cat chipotle.tsv | grep "Chicken Burrito" chipotle.tsv | grep "Black Beans" | wc -l

282

danielles-air:data Danielle$ cat chipotle.tsv | grep "Chicken Burrito" chipotle.tsv | grep "Pinto Beans" | wc -l

105

More have black beans

1. Make a list of all of the CSV or TSV files in the [our class repo] (https://github.com/ga-students/DS-SEA-3). repo (using a single command). You will be working on your local repo on your laptop. Think about how wildcard characters can help you with this task.

danielles-air:DS-SEA-06 Danielle$ find data /\*.csv | find data /\*tsv | wc -l

find: /\*.csv: No such file or directory

find: /\*tsv: No such file or directory

48

1. Count the approximate number of occurrences of the word "dictionary" (regardless of case) across all files of [our class repo] (<https://github.com/ga-students/DS-SEA-3>).

danielles-air:DS-SEA-06 Danielle$ grep -r -i "dictionary" . | wc -w

1174

1. Optional: Use the the command line to discover something "interesting" about the Chipotle data. Try using the commands from the "advanced" section!

Finding unique lines:

danielles-air:data Danielle$ sort chipotle

...

danielles-air:data Danielle$ uniq -c chipotle.tsv | wc -l

4589

danielles-air:data Danielle$

Out of 4623