What are WIC households' total expenditures on whole wheat bread?

Import libraries and connect to database

We will want to join household level data to purchasing data, which is in the `trip_all` table. Let's read that data in.

In [ ]:
```
# pandas-related imports
import pandas as pd

# database interaction imports
from pyathenajdbc import connect
```

In [ ]:
```
conn = connect(s3_staging_dir = 's3://usda-iri-2019-queryresults/',
               region_name = 'us-gov-west-1',
               LogLevel = '0',
               workgroup = 'workgroup-iri_usda')
```

In [ ]:
```
bread_query = """
SELECT distinct upc, flavor,upcdesc
from iri_usda.pd_pos_all
where upcdesc like '%100%'
and product in ('FRESH BREAD','HAMBURGER AND HOT DOG BUNS','PITA BREA
D','BAGELS/BIALYS','BREAD','ROLL','BUN'
    ,'BAGEL')
and category = 'FRESH BREAD & ROLLS';"""

bread_df = pd.read_sql(bread_query, conn)
```

In [ ]:
```
ww_flavor_terms = ['WHOLE','WHEAT','WHOLE WHEAT','GRAIN','OAT']
ww_df = bread_df[bread_df.flavor.str.contains('|'.join(ww_flavor_terms
))]
```

Now we get the UPC codes.

In [ ]:
```
ww_upc_list = ww_df.upc.unique().tolist()
len(ww_upc_list)
```

In [ ]:
```
wic_hh_query = """
SELECT distinct panid
from iri_usda_2019_db.project_cohort
where projection61k > 0 and wic_june = 1;"""
wic_hh_df = pd.read_sql(wic_hh_query, conn)
```

In [ ]:
```
wic_hh_list = wic_hh_df.panid.unique().tolist()
```

In [ ]:
```
trip_query = """
SELECT distinct purdate,panid,mop,upc,dollarspaid
from iri_usda.trip_all
where year = '2017' and
panid in {} and
upc in {};""".format(tuple(wic_hh_list),tuple(ww_upc_list))
```

In [ ]:
```
bread_trip_df = pd.read_sql(trip_query,conn)
```

In [ ]:
```
joined = pd.merge(wic_hh_df,bread_trip_df, on = 'panid')
```

In [ ]:
```
wic_purchases = joined.loc[joined.mop == '7']
wic_purchases.shape
```

In [ ]:
```
len(wic_purchases.panid.unique().tolist())
```

In [ ]:
```
wic_purchases['month'] = wic_purchases['purdate'].apply(lambda x: x.mont
h)
```

In [ ]:
```
wic_purchases_sub = wic_purchases[['month','dollarspaid']]
```

In [ ]:
```
wic_purchases_agg = wic_purchases_sub.groupby(['month']).sum()
wic_purchases_agg
```

In [ ]:
```
Import visualization packages
import matplotlib.pyplot as plt
import seaborn as sn

# so images get plotted in the notebook
%matplotlib inline
```

In [ ]:
```
ax = wic_purchases_agg.plot(figsize = (12, 6))
ax.set(ylabel = '$ Spent', title = '$ Spent by WIC-Households on 100% Wh
ole Wheat Bread by month, 2017')
ax.get_legend().remove()
plt.annotate('Sources: IRI Consumer Network and InfoScan',
             xy=(0.75,-0.1));
```