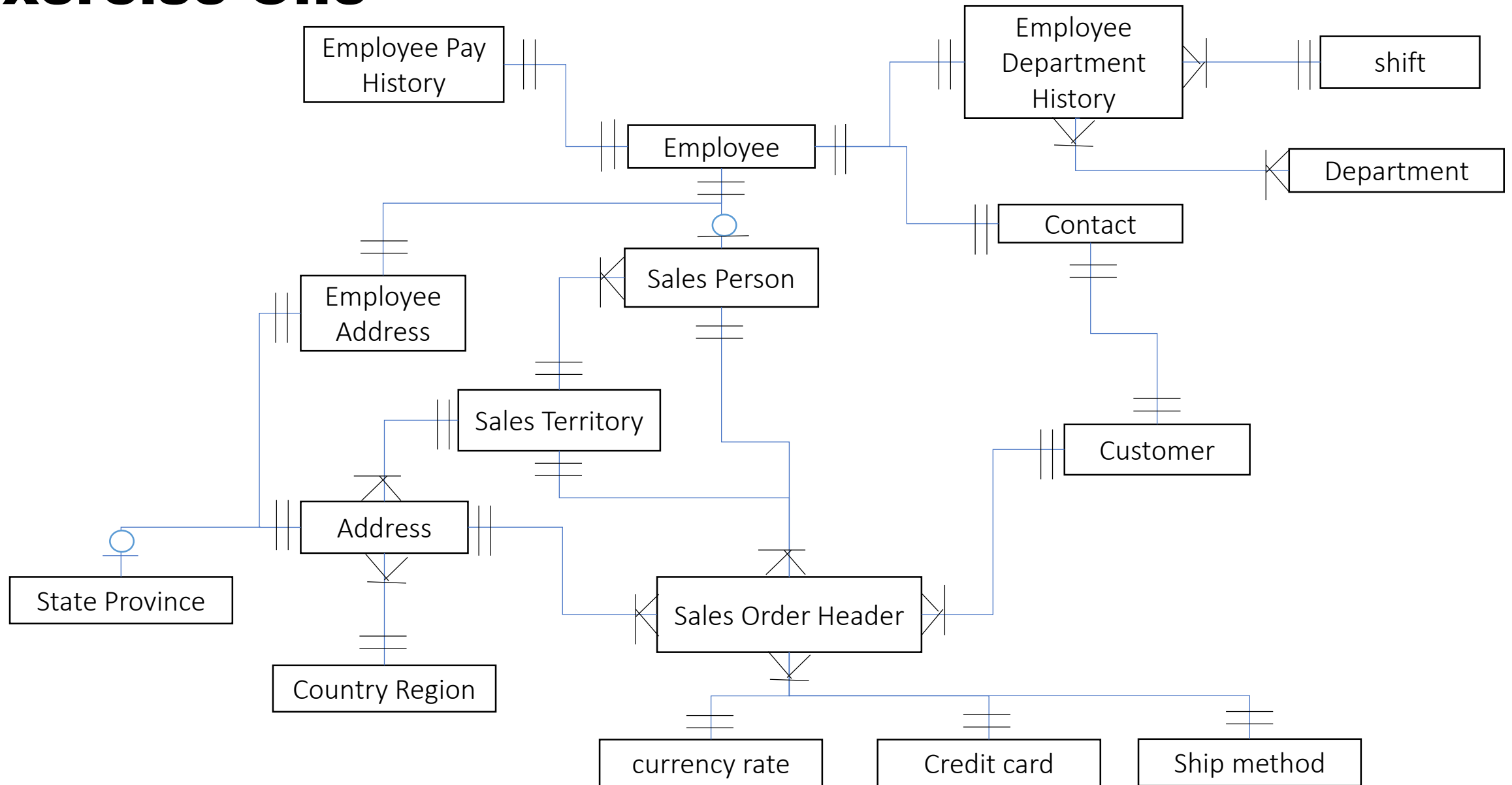# Reporting Challenge

Miya Wang

7/11/2017
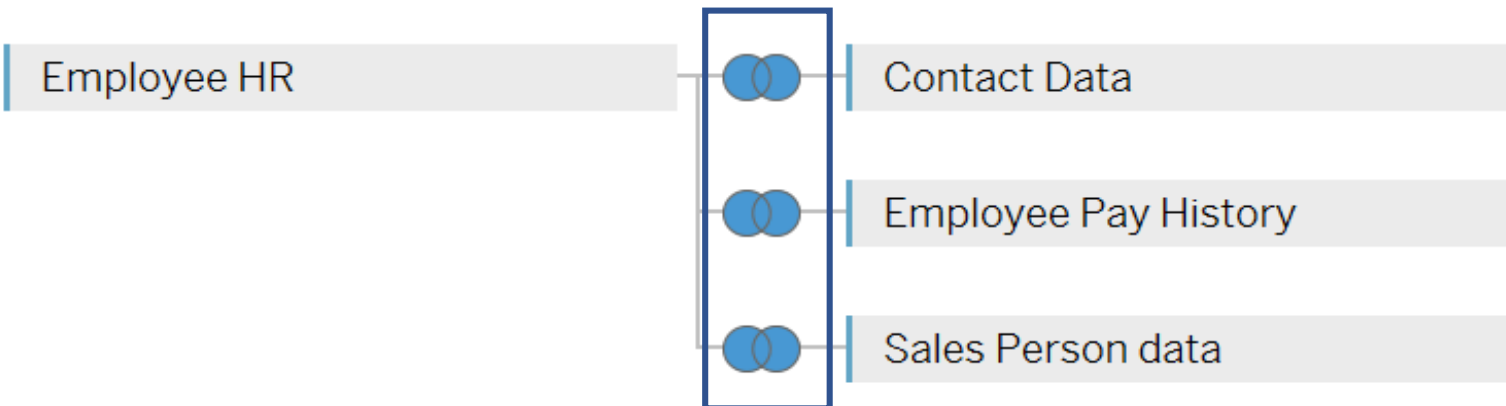
Exercise One

Exercise Two

```python
dfmerged = employeeHR.merge(employeePay,on = 'Business Entity ID',
                    how='outer'
            ).merge(contact,on = 'Business Entity ID',
                    how='outer'
            ).merge(salesperson,on = 'Business Entity ID',
                    how='outer')
dfmerged = dfmerged.reindex_axis(sorted(dfmerged.columns), axis=1)
dfmerged.shape
```

(290, 41)

| Employee HR | | Contact Data |
| Employee Pay History |
| Sales Person data |

- *Full Outer Join of three spreadsheets*
- *Result in a dataset with 290 records and 41 features*

Exercise Three

```
df_1 = dfmerged[['Business Entity ID', "Last Name", "First Name",
                 "Job Title", "Rate", "Bonus", "Sales Quota",
                 "Sales Last Year"]].sort_values(by = 'Rate',ascending = False)
df_1['Name'] = df_1["First Name"]+' '+df_1["Last Name"]
df_1.drop("First Name",axis = 1).drop("Last Name",axis = 1).head()
```

	Business Entity ID	Job Title	Rate	Bonus	Sales Quota	Sales Last Year	Name
0	1.0	Chief Executive Officer	125.5000	NaN	NaN	NaN	Ken Sánchez
26	25.0	Vice President of Production	84.1346	NaN	NaN	NaN	James Hamilton
288	273.0	Vice President of Sales	72.1154	NaN	NaN	NaN	Brian Welcker
1	2.0	Vice President of Engineering	63.4615	NaN	NaN	NaN	Terri Duffy
248	234.0	NaN	60.0962	NaN	NaN	NaN	Laura Norman

- *Sort by Salary Rate*
- *Total the Columns*
- *Merge First Name and Last Name column into One*

Exercise Three

1. Who has the paid the highest salary rate in the firm? What is their job title?

```
df_1.sort_values(by = 'Rate',ascending = False).ix[0][['Name','Job Title']]
```

```
Name                    Ken Sánchez
Job Title      Chief Executive Officer
```

2. Who got the highest Bonus last year? How much?

```
print(df_1.sort_values(by = 'Bonus',ascending = False).dropna(
        subset = ['Bonus'])[:1][['Name','Bonus']])
```

```
            Name     Bonus
300   Tsvi Reiter    6700.0
```

3. Which salesperson had the highest sales last year? How much?

```
print(df_1.sort_values(by = 'Sales Last Year',ascending = False).dropna(
        subset = ['Sales Last Year'])[:1][['Name','Sales Last Year']])
```

```
                  Name    Sales Last Year
322   Ranjit Varkey Chudukatil      2.396540e+06
```

4. Did any of our salespeople not meet their quota? If so, who was it?

```
df_1_a = df_1.dropna(subset = ['Sales Last Year','Sales Quota'])
print(df_1_a[df_1_a['Sales Last Year']<df_1_a['Sales Quota']][['Name']])
```

```
            Name
310   Tete Mensa-Annan
```

Exercise Three

5. What is the mean rate of pay? Median? Mode?

```
df_1['Rate'].dropna().describe()
```

```
count     290.000000
mean       18.188257
std        12.290579
min         9.000000
25%        11.000000
50%        14.000000
75%        23.076900
max       125.500000
Name: Rate, dtype: float64
```

```
from collections import Counter
Counter(df_1['Rate'].dropna()).most_common(1)
```

```
[(9.5, 27)]
```

6. What other interesting insights can be derived?

```
Counter(df_1['Job Title'].dropna()).most_common(10)
```
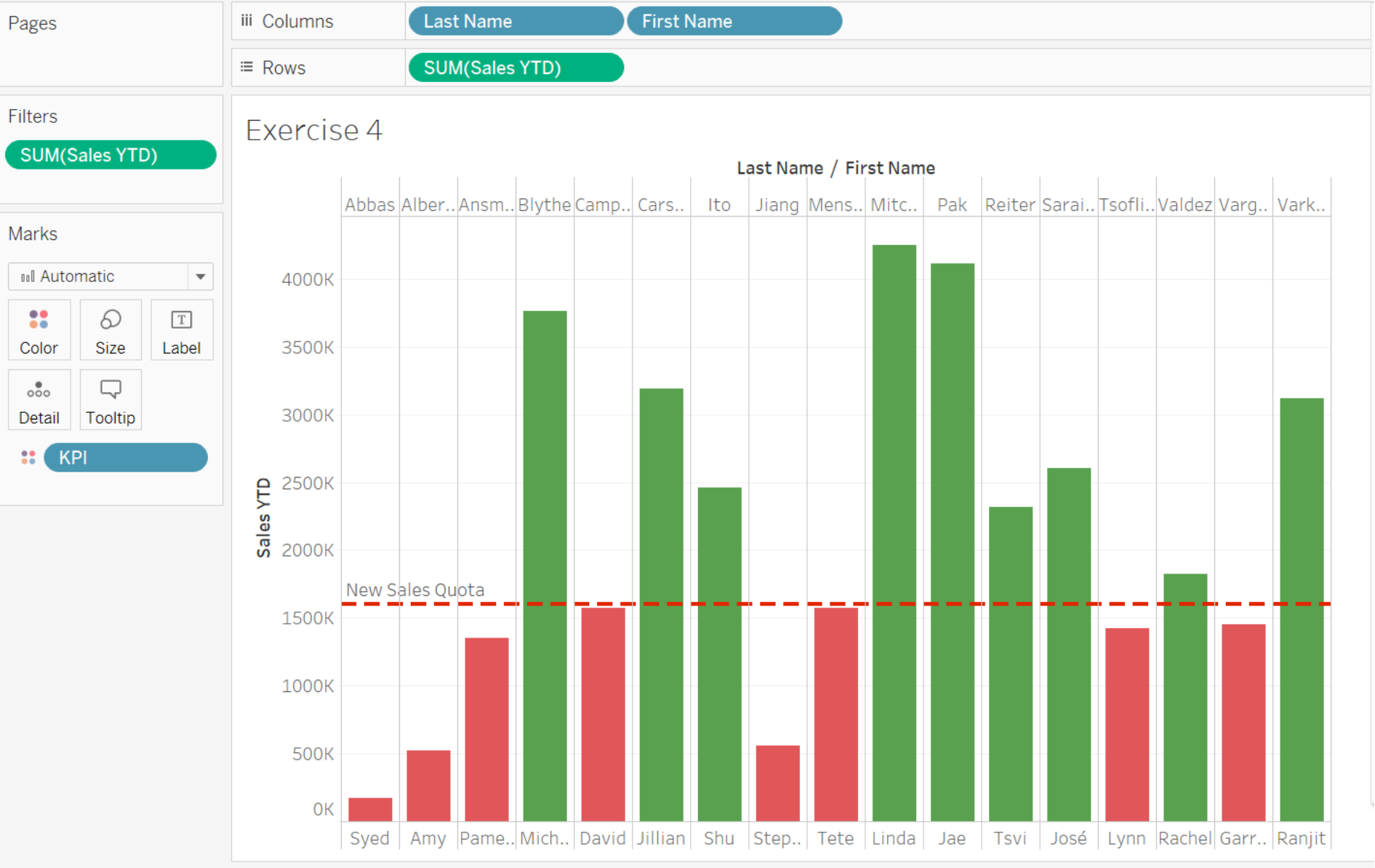
```
[('Production Technician - WC40', 26),
 ('Production Technician - WC50', 26),
 ('Production Technician - WC60', 26),
 ('Production Technician - WC30', 25),
 ('Production Technician - WC20', 22),
 ('Production Technician - WC10', 17),
 ('Production Technician - WC45', 15),
 ('Sales Representative', 14),
 ('Buyer', 9),
 ('Marketing Specialist', 5)]
```

```
df_1_b  = df_1.dropna(subset = ['Job Title'])
Counter(df_1_b[df_1_b['Job \
Title'].str.contains("Production \
Technician")]['Rate'].dropna()).most_common()
```

```
[(11.0, 26),
 (12.449999999999999, 26),
 (15.0, 26),
 (14.0, 22),
 (9.5, 17),
 (13.449999999999999, 17),
 (10.0, 15)]
```
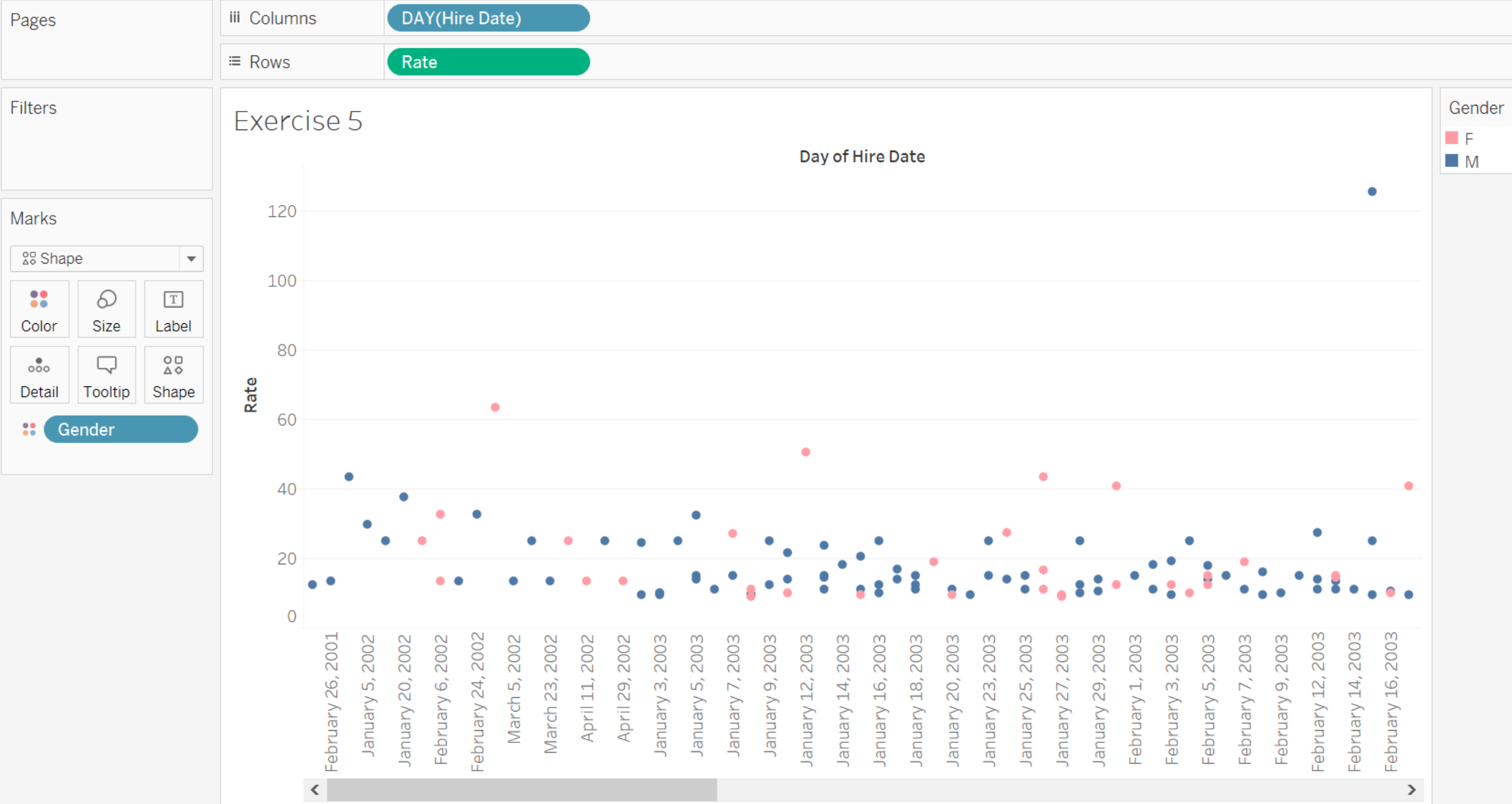
- *The most common position in company is Product Technician*
- *Product technician is usually paid at the rate of 11, 12.5 or 15*
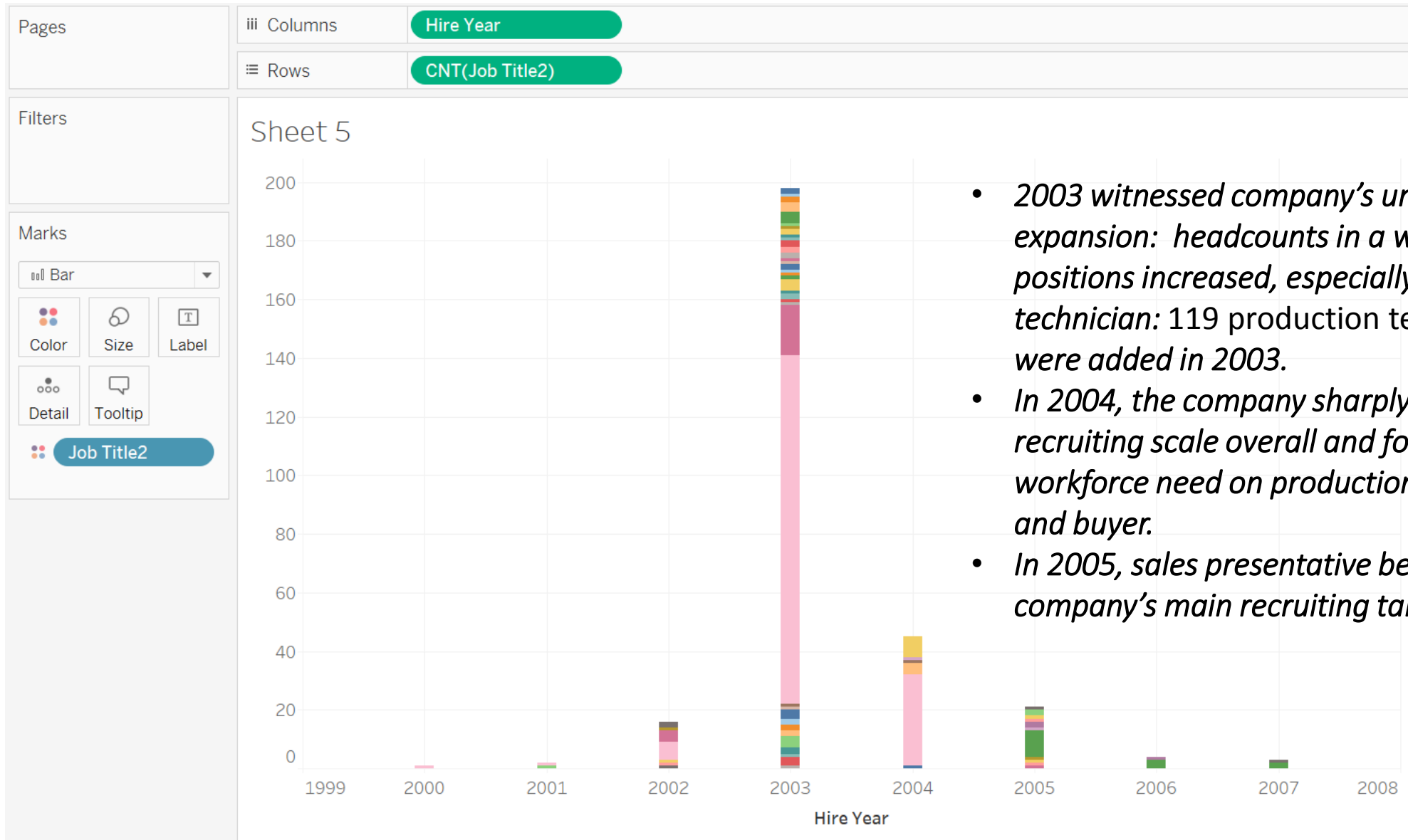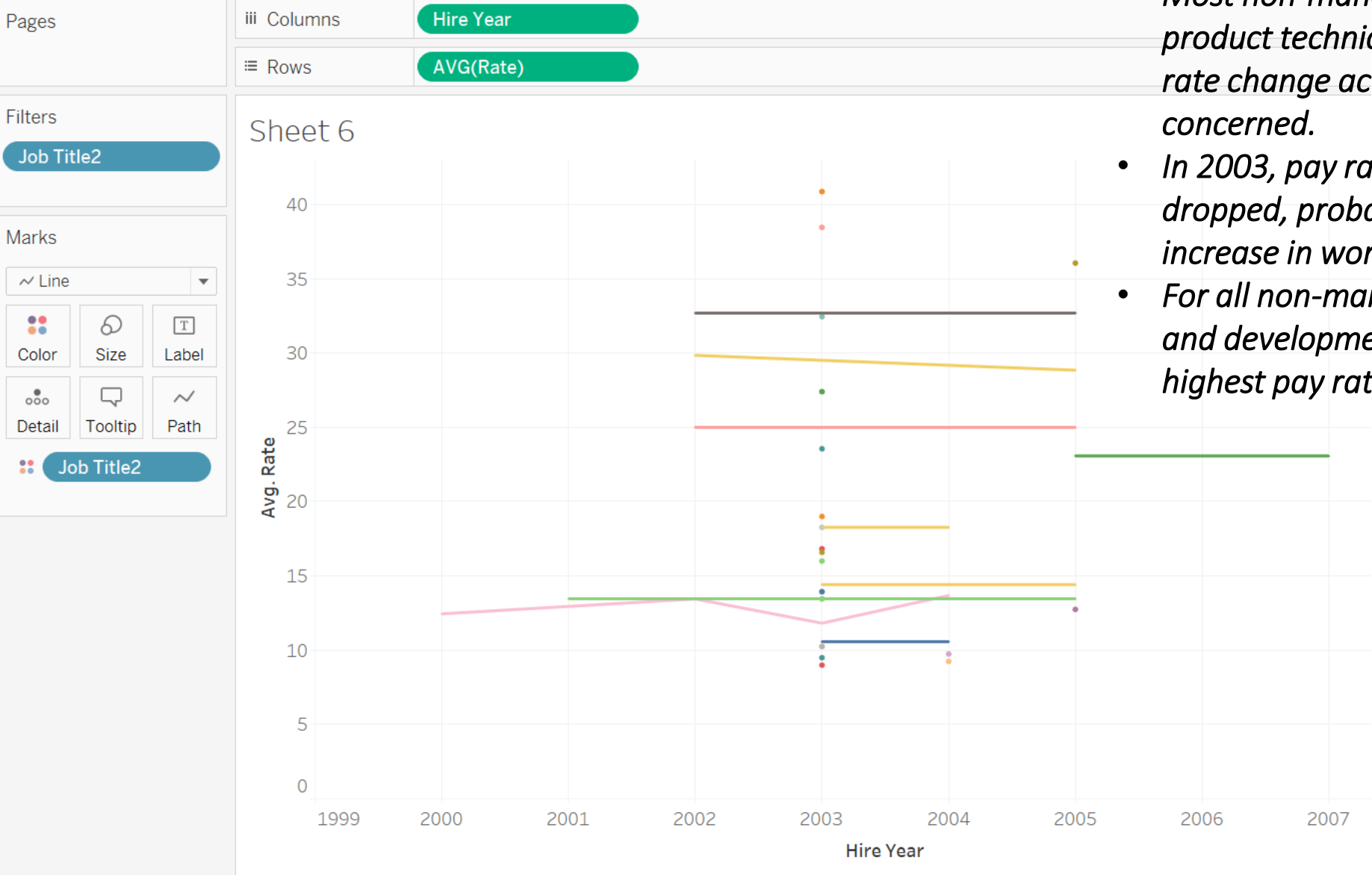
6

Exercise Four

Exercise Five

Exercise Six

- *2003 witnessed company's unprecedented expansion: headcounts in a wide range of positions increased, especially production technician:* 119 production technician *were added in 2003.*
- *In 2004, the company sharply shrunk its recruiting scale overall and focused its workforce need on production technician and buyer.*
- *In 2005, sales presentative became the company's main recruiting target.*

Exercise Six

- *Most non-managerial positions, except product technician, didn't experience pay rate change across the time period concerned.*
- *In 2003, pay rate for product technician dropped, probably due to the large increase in workforce in the same year.*
- *For all non-managerial position, research and development engineer enjoyed the highest pay rate: 40.87 on average.*