Interim Project
Adventure works 2019

Ideas, sketches and key components

by group 4 art

1. What are the regional sales in the best performing country?

Sales.CountryRegionCurrency?

Sales.SalesTerritory?

Sales.SalesOrderHeader?

Sale.OrderDetail?

Will probably need to sum the order amounts as Sales.SalesTerritory doesn't have all the countries listed.

Where to find regional sales? Or need to create a view/join tables?

1. Find best performing country

2. Find the total for regional sales in country.

Person.countryregion
Person.StateProvince
Sales.Territory
Sales.Customer - counted numb within each territory. Confirmed
Sales.Territory contained all territories for sales.
Summed SalesYTD and found total tales for each country, sorted in desc.

Not enough info to tell us what sales territory meant, so we had to find the bottom most table.

2. What is the relationship between annual leave taken and bonus?

[HumanResources].[Employee]

[VacationHours]
[SickLeaveHours]

[AdventureWorks].[Sales].[SalesPerson]

[Bonus]

BusinessEntityID	Bonus	JobTitle	VacationHours
274	0.00	North American Sales Manager	14
275	4100.00	Sales Representative	38
276	2000.00	Sales Representative	27
277	2500.00	Sales Representative	24
278	500.00	Sales Representative	33
279	6700.00	Sales Representative	29
280	5000.00	Sales Representative	22
281	3550.00	Sales Representative	26
282	5000.00	Sales Representative	31
283	3500.00	Sales Representative	23
284	3900.00	Sales Representative	39
285	0.00	Pacific Sales Manager	20
286	5650.00	Sales Representative	36
287	0.00	European Sales Manager	21
288	75.00	Sales Representative	35
289	5150.00	Sales Representative	37
290	985.00	Sales Representative	34

Include, exclude managers?

Scatter plot

Compare sales team with and without managers as different correlation

3 managers – no sales, no bonus! exclude or include?

Why did 284 got a bonus? No sales for last year!

2. What is the relationship between annual leave taken and bonus?

[HumanResources].[Employee]

[VacationHours]
[SickLeaveHours]

Some employees have more than one pay entry. Bonus? Or changing departments?

What connects them? Date?

How to discard 250

BusinessEntityID	RateChangeDate	Rate
4	2007-12-05 00:00:00.000	8.62
4	2010-05-31 00:00:00.000	23.72
4	2011-12-15 00:00:00.000	29.8462

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
4	1	1	2007-12-05	2010-05-30
4	2	1	2010-05-31	

BusinessEntityID	RateChangeDate	Rate
16	2007-12-20 00:00:00.000	24.00
16	2009-07-15 00:00:00.000	28.75
16	2012-04-30 00:00:00.000	37.50

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
16	5	1	2007-12-20	2009-07-14
16	4	1	2009-07-15	

BusinessEntityID	RateChangeDate	Rate
167	2008-12-02 00:00:00.000	6.50
167	2011-12-01 00:00:00.000	7.25
167	2013-07-14 00:00:00.000	9.50

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
167	7	1	2008-12-02	NULL

BusinessEntityID	RateChangeDate	Rate
170	2008-12-21 00:00:00.000	6.50
170	2011-12-01 00:00:00.000	7.25
170	2013-07-14 00:00:00.000	9.50

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
170	7	1	2008-12-21	NULL

BusinessEntityID	RateChangeDate	Rate
172	2009-01-16 00:00:00.000	6.50
172	2011-12-01 00:00:00.000	7.25
172	2013-07-14 00:00:00.000	9.50

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
172	7	1	2009-01-16	NULL

BusinessEntityID	RateChangeDate	Rate
174	2009-01-27 00:00:00.000	6.75
174	2011-12-01 00:00:00.000	7.25
174	2013-07-14 00:00:00.000	9.50
175	2009-02-02 00:00:00.000	6.50
175	2011-12-01 00:00:00.000	7.25
175	2013-07-14 00:00:00.000	9.50
176	2009-02-14 00:00:00.000	6.50
176	2011-12-01 00:00:00.000	7.25
176	2013-07-14 00:00:00.000	9.50
177	2009-02-21 00:00:00.000	6.50
177	2011-12-01 00:00:00.000	7.25
177	2013-07-14 00:00:00.000	9.50
178	2009-03-03 00:00:00.000	6.50
178	2011-12-01 00:00:00.000	7.00
178	2013-07-14 00:00:00.000	9.50

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
174	7	1	2009-01-27	NULL
175	7	1	2009-02-02	NULL
176	7	1	2009-02-14	NULL
177	7	1	2009-02-21	NULL
178	7	1	2009-03-03	NULL

BusinessEntityID	RateChangeDate	Rate
224	2008-01-08 00:00:00.000	8.62
224	2011-08-31 00:00:00.000	13.50
224	2012-01-01 00:00:00.000	16.00

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
224	7	1	2009-01-07	2011-08-31
224	8	1	2011-09-01	NULL

BusinessEntityID	RateChangeDate	Rate
234	2009-01-31 00:00:00.000	39.06
234	2011-11-14 00:00:00.000	48.5577
234	2012-01-29 00:00:00.000	60.0962

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
234	10	1	2009-01-31	2013-11-13
234	16	1	2013-11-14	NULL

BusinessEntityID	RateChangeDate	Rate	PayFrequency	ModifiedDate
250	2011-02-21 00:00:00.000	9.862		2011-02-07 00:00:00.000
250	2011-07-30 00:00:00.000	22.50	2	2011-07-16 00:00:00.000
250	2012-07-14 00:00:00.000	30.00	2	2012-06-30 00:00:00.000

BusinessEntityID	DepartmentID	ShiftID	StartDate	EndDate
250	4	1	2011-02-25	2011-07-30
250	13	1	2011-07-31	2012-07-14
250	5	1	2012-07-15	NULL

250 changed departments twice

3. What is the relationship between Country and Revenue?

Proportion of revenue per country?

Group country using SQL then visualise in Python

Bar graph

Speciality -
Sales.vStoreWithDemographics
Country

1. bar chart showing country(x) and revenue(y)	**2. Number of stores per country**	**Number of sales per country?**

4. What is the relationship between sick leave and Job Title (PersonType)?

[HumanResources].[Employee]

[SickLeaveHours]

[JobTitle]

**How to group?
Job title?
Organisational level?
Department?**

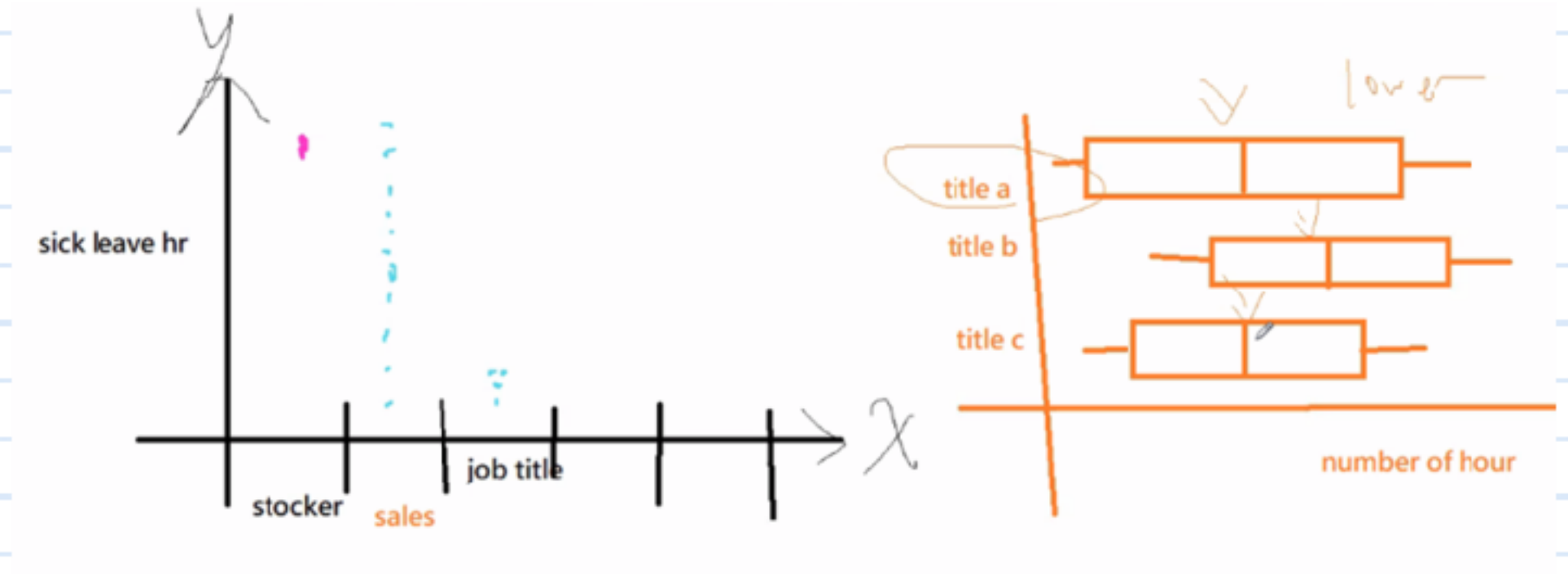**Box plots?**

Job title – 67 categories – too big

Averages? not representative

Org Level - too small
NULL   0   69
1   6   209
2   27   1224
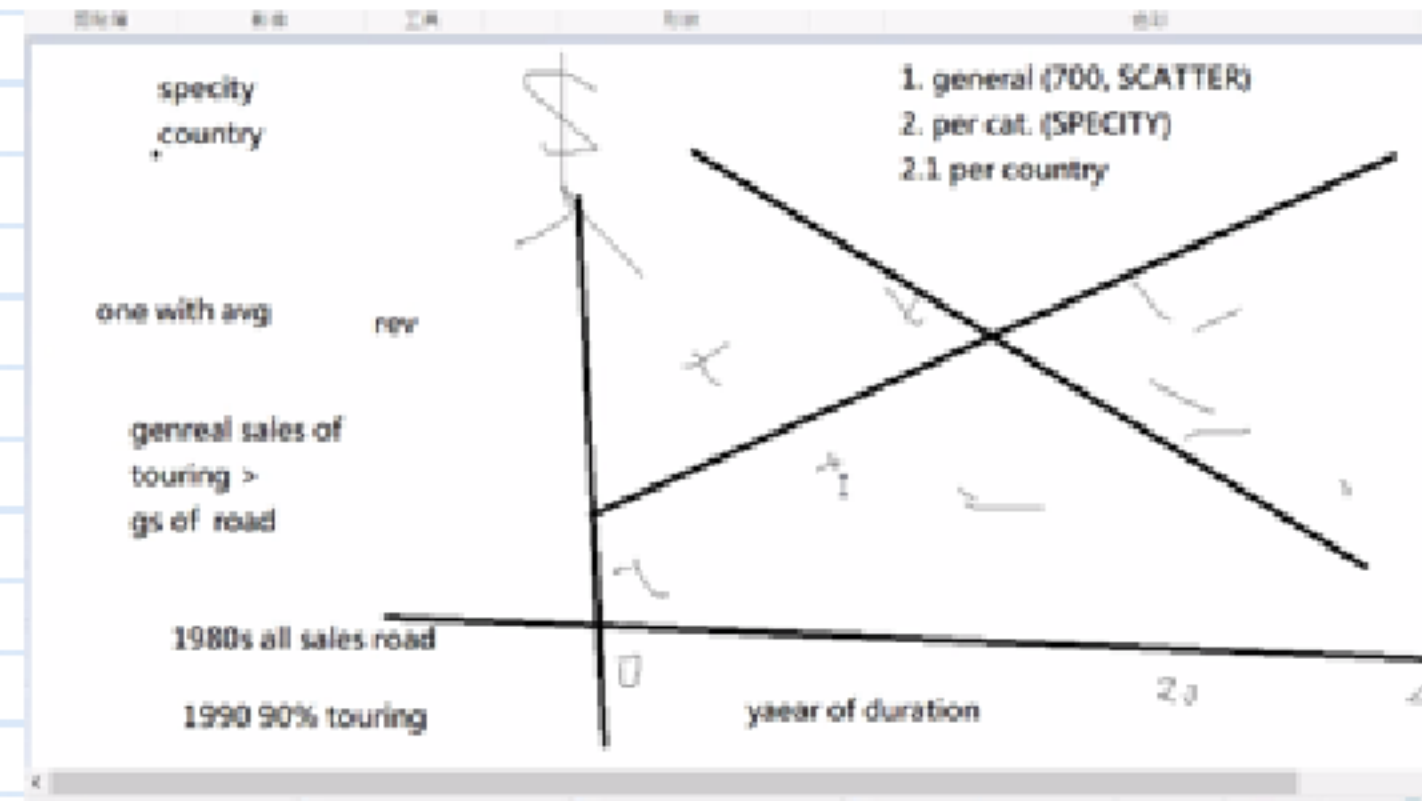3   66   3088
4   190 8549

Department – 16 categories

5. What is the relationship between store trading duration and revenue?

HumanResources.Shift?

store.demo graphics

Compare scatter plots and histograms

Averages?


```
specity
country

one with avg          rev

genreal sales of
touring >
gs of road

1980s all sales road

1990 90% touring
```

```
1. general (700, SCATTER)
2. per cat. (SPECITY)
2.1 per country
```

yaear of duration

```sql
SELECT  2014 - [YearOpened] AS Store_Trading_Duration,
AVG([AnnualRevenue]) AS Average_Revenue
FROM [AdventureWorks].[Sales].[vStoreWithDemographics]
GROUP BY 2014 - [YearOpened]
```

6. What is the relationship between the size of the stores, number of employees and revenue? Sales.store?

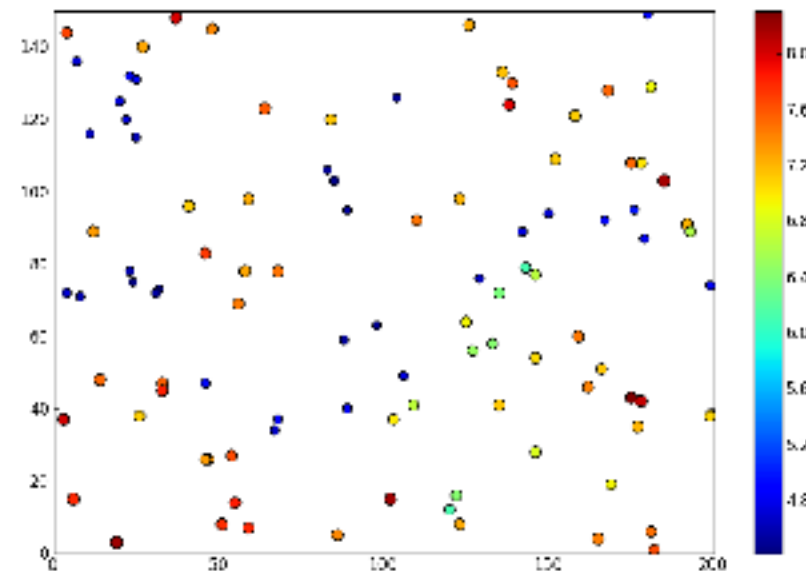| stores vs number of employees | revenue vs number of employees | stores vs revenue |

3 dimensional scatter plot

Check with color, size or transparency

Correlation coefficient

Viridis or design our own color scale?


```
SELECT [BusinessEntityID]
      ,[AnnualRevenue]
      ,[SquareFeet]
      ,[NumberEmployees]
  FROM [AdventureWorks].[Sales].
[vStoreWithDemographics]"' # string is the sql query
```

Add 2D graphs or go with 3D only?