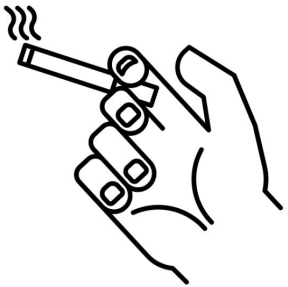
Birth Registration AND Potential Influential Variables

2022 AU INFO499 Group 1

Problem Space

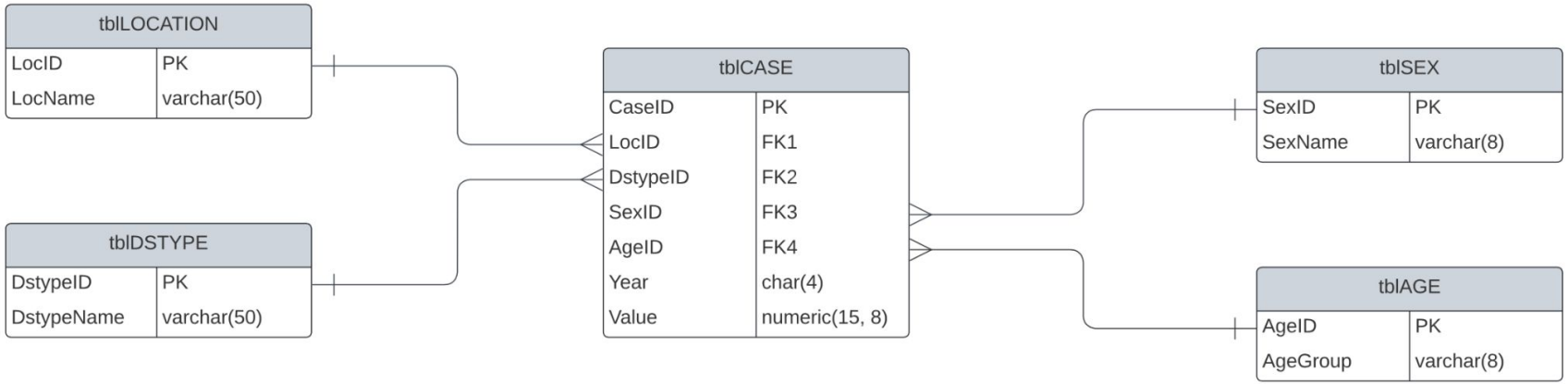
What impact does the use of tobacco, particularly by chewing, and the prevalence of obesity and overweight have on fertility rates and/or population growth worldwide from 1990 to 2015?



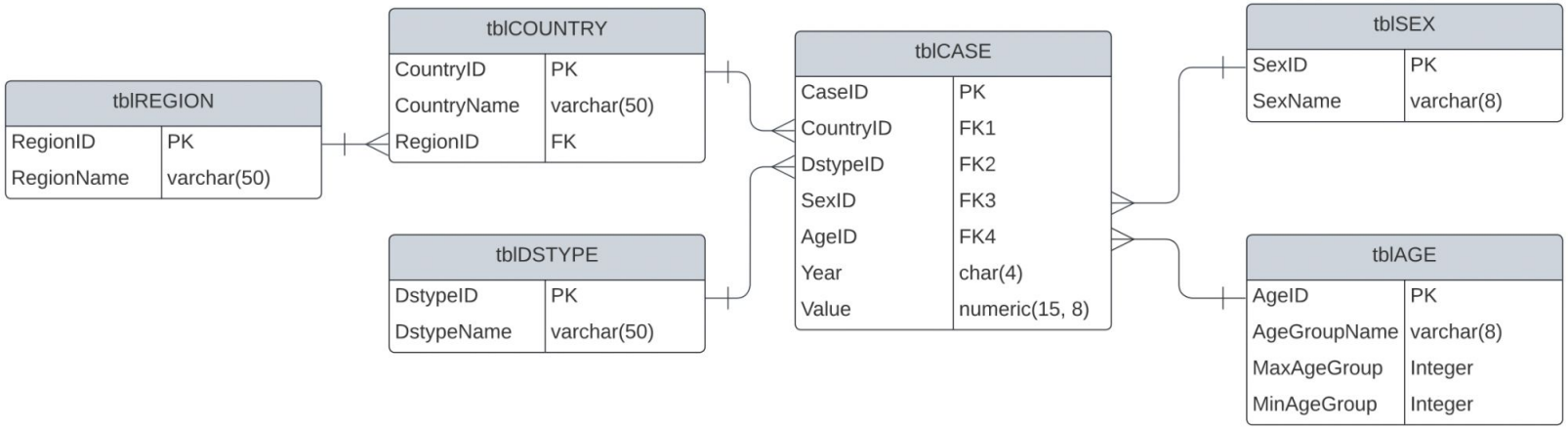
01

Our ERD

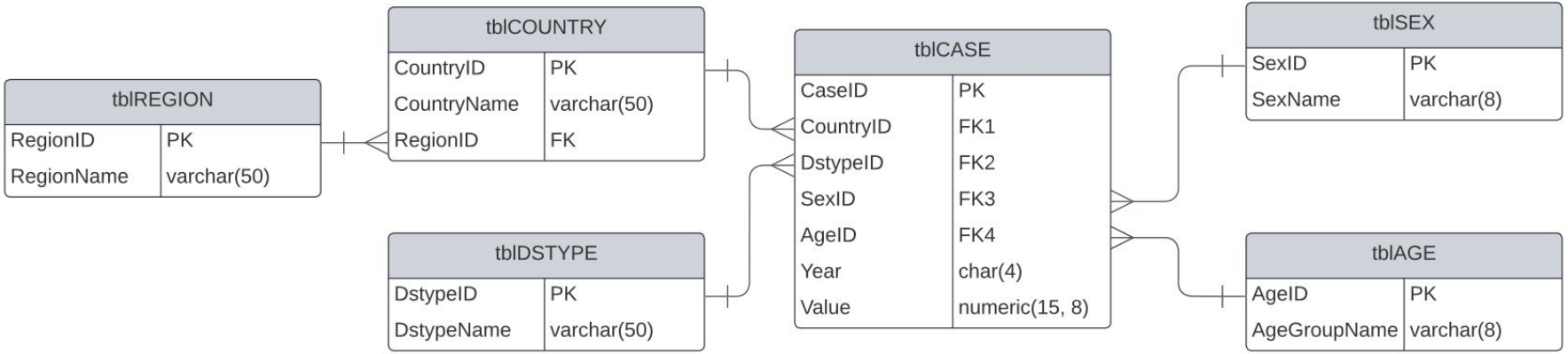
Draft.ERD


tblLOCATION
LocID
PK
LocName
varchar(50)
tblDSTYPE
DstypeID
PK
DstypeName
varchar(50)
tblCASE
CaseID
PK
LocID
FK1
DstypeID
FK2
SexID
FK3
AgeID
FK4
Year
char(4)
Value
numeric(15, 8)
tblSEX
SexID
PK
SexName
varchar(8)
tblAGE
AgeID
PK
AgeGroup
varchar(8)

2nd Draft ERD



Final ERD


tblREGION
RegionID PK
RegionName varchar(50)
tblCOUNTRY
CountryID PK
CountryName varchar(50)
RegionID FK
tblDSTYPE
DstypeID PK
DstypeName varchar(50)
tblCASE
CaseID PK
CountryID FK1
DstypeID FK2
SexID FK3
AgeID FK4
Year char(4)
Value numeric(15, 8)
tblSEX
SexID PK
SexName varchar(8)
tblAGE
AgeID PK
AgeGroupName varchar(8)

DDL/Physical Database Design

01

Pre-processing

Raw data cleaning that simplifies the insertion process.

02

Populating

Complex Wrapper transactions to insert 5,000,000+ rows of data

03

Computed Columns

Advanced calculation of existing raw columns

04

Business Rule

Constraints that avoids impractical data insertions

05

Views

Complex queries to filter data

Pre-processing

- Python Pandas
  - Filtering out the needed columns
  - Concatenating datasets to a single big one

```
tobacco_data.head()
```

	location_name	sex_name	age_group_name	ds_type
1710	China	Male	15 to 19	Chewing t
1711	China	Female	15 to 19	Chewing t
1712	China	Male	15 to 19	Chewing t
1713	China	Female	15 to 19	Chewing t
1714	China	Male	15 to 19	Chewing t

```
obesity.head()
```

	location_name	sex_name	age_group_name	year	
648	China	Male	5 to 9	1980	
649	China	Male	5 to 9	1980	4958
650	China	Female	5 to 9	1980	
651	China	Female	5 to 9	1980	4703
652	China	Both	5 to 9	1980	

```
overweight.head()
```

	location_name	sex_name	age_group_name	year
648	China	Male	5 to 9	1980
649	China	Male	5 to 9	1980
650	China	Female	5 to 9	1980
651	China	Female	5 to 9	1980
652	China	Both	5 to 9	1980

```
violence.head()
```

	location_name	year	ds_type_name	val
0	China	1990	Int Partner Viol	8.277482
1	China	1991	Int Partner Viol	8.242396
2	China	1992	Int Partner Viol	8.205062
3	China	1993	Int Partner Viol	8.166232
4	China	1994	Int Partner Viol	8.127625

```
population.head()
```

	ds_type_name	location_name	year	sex_name	age_group_name	val
864	Population	China	1950	male	<1 year	9.251046e+06
865	Population	China	1950	male	1	9.753580e+06
866	Population	China	1950	male	2	7.722766e+06
867	Population	China	1950	male	3	7.688385e+06
868	Population	China	1950	male	4	7.265009e+06

```
birth.head()
```

	location_name	year	age_group_name	val
0	Albania	1948	10 to 14	0.0
1	Albania	1948	15 to 19	1743.0
2	Albania	1948	20 to 24	10013.0
3	Albania	1948	25 to 29	11778.0
4	Albania	1948	30 to 34	8214.0

Populating

- Basic one-row inserting stored procedures
- Populating stored procedures for lookup tables & child tables
- “GetId” multiple nested stored procedures and explicit transactions to manage INSERT activity
- Wrapper algorithms to execute mass insertions with throughput volume of 5,000,000 rows data

```
90   SELECT COUNT(*) FROM tblCASE
```

Results | Messages

	(No column name)
1	5872547

```
INSERT INTO tblREGION (RegionName) SELECT distinct region FROM IHME_raw_data WHERE region IS NOT NULL
SELECT*FROM tblREGION

INSERT INTO tblDSTYPE (DstypeName) SELECT distinct ds_type_name FROM IHME_raw_data WHERE ds_type_name IS NOT NULL
SELECT*FROM tblDSTYPE

INSERT INTO tblSEX (SexName) SELECT distinct sex_name FROM IHME_raw_data WHERE sex_name IS NOT NULL
SELECT*FROM tblSEX

INSERT INTO tblAGE (AgeGroupName) SELECT distinct age_group_name FROM IHME_raw_data WHERE age_group_name IS NOT NULL
SELECT*FROM tblAGE
```

```
CREATE PROCEDURE GetRegionID
@RN varchar(50),
@Region_ID INT OUTPUT
AS
SET @Region_ID = (SELECT RegionID FROM tblREGION WHERE RegionName = @RN)
GO

CREATE PROCEDURE GetCountryID
@CN varchar(50),
@Country_ID INT OUTPUT
AS
SET @Country_ID = (SELECT CountryID FROM tblCOUNTRY WHERE CountryName = @CN)
GO

CREATE PROCEDURE GetDstypeID
```

```
55  CREATE OR ALTER PROCEDURE wrapper_INSERT_Case
56  AS
57
58  DECLARE @CountryName varchar(50)
59  DECLARE @Dstype varchar(50)
60  DECLARE @Sex varchar(50)
61  DECLARE @Age varchar(50)
62  DECLARE @CaseYear char(4)
63  DECLARE @CaseValue float
64  DECLARE @RUN INT
65  SET @RUN = (SELECT COUNT(*) FROM ##tblCASE)
66
67  WHILE @Run > 0
68  BEGIN
69
70  SET @CountryName = (SELECT location_name FROM ##tblCASE WHERE @Run = PK)
71  SET @Dstype = (SELECT ds_type_name FROM ##tblCASE WHERE @Run = PK)
72  SET @Sex = (SELECT sex_name FROM ##tblCASE WHERE @Run = PK)
73  SET @Age = (SELECT age_group_name FROM ##tblCASE WHERE @Run = PK)
74  SET @CaseYear = (SELECT [year] FROM ##tblCASE WHERE @Run = PK)
75  SET @CaseValue = (SELECT val FROM ##tblCASE WHERE @Run = PK)
76
77  EXEC INSERT_Case
78  @CN2 = @CountryName,
79  @DSTN2 = @Dstype,
80  @SN2 = @Sex,
81  @AGN2 = @Age,
82  @Year = @CaseYear,
83  @Value = @CaseValue
```

Computed Column

The computed columns we made:

- the ratio of the population and the birth registration in each country
- counts country numbers in different regions

Computed Column

```sql
--- Computed Column: a column of the ratio of population and birth registration in each country
CREATE FUNCTION fn_RatioPopulationBirth(@PK INT)
RETURNS NUMERIC(10,2)
AS
BEGIN

DECLARE @RET NUMERIC(10,2) = (SELECT A.CaseValue/B.CaseValue AS RatioPopulationBirth
                              FROM
                              (SELECT C.CaseValue, CO.CountryID
                              FROM tblCase C
                                  JOIN tblDsType DT ON C.DsTypeID = DT.DsTypeID
                                  JOIN tblCountry CO ON C.CountryID = CO.CountryID
                              WHERE DT.DstypeName = 'Birth Regitration')A,

                              (SELECT C.CaseValue, CO.CountryID
                              FROM tblCase C
                                  JOIN tblDsType DT ON C.DsTypeID = DT.DsTypeID
                                  JOIN tblCountry CO ON C.CountryID = CO.CountryID
                              WHERE DT.DstypeName = 'Population')B
                              WHERE A.CountryID = B.CountryID
                                  AND A.CountryID = @PK)
RETURN @RET
END
GO

ALTER TABLE tblGENRE
ADD NumMovie AS (dbo.fn_NumMovieEachGenre(GenreID))
GO
```

```sql
--- Computed Column: a column that count country numbers in different region
CREATE FUNCTION fn_CountryInRegion(@PK INT)
RETURNS NUMERIC(10,2)
AS
BEGIN

DECLARE @RET NUMERIC(10,2) = (SELECT COUNT(DISTINCT CO.CountryID) AS NumCountry
                              FROM tblCountry CO
                                  JOIN tblRegion R ON CO.RegionID = R.RegionID
                              WHERE R.RegionID = @PK)
RETURN @RET
END
GO

ALTER TABLE tblRegion
ADD NumCountry AS (dbo.fn_CountryInRegion(RegionID))
GO
```

Business Rules

No **Male** in the age group of **30 to 34** in **Japan** should chew tobacco because it's bad for fertility.

```
CREATE FUNCTION No_Male_30_34_Japan()
RETURNS INT
AS
BEGIN

DECLARE @RET INT = 0
IF EXISTS(SELECT * FROM tblCase C
JOIN tblCountry CO ON C.CountryID = CO.CountryID
JOIN tblSEX S ON C.SexID = S.SexID
JOIN tblAGE A ON C.AgeID = A.AgeID
JOIN tblDSTYPE DT ON C.DstypeID = DT.DstypeID
WHERE CO.CountryName = 'Japan'
AND S.SexName = 'Male'
AND A.AgeGroupName = '30 to 34'
AND C.CaseValue= 0)

SET @RET = 1

RETURN @RET
END
GO

ALTER TABLE tblCASE WITH NOCHECK
ADD CONSTRAINT RestrictAgeCountry
CHECK (dbo.No_Male_30_34_Japan() = 0)
GO
```

Business Rules

Age group 15 to 19 cannot chew tobacco in China because it is illegal

```
CREATE FUNCTION AgeGroup_Tobacco_Rule()
RETURNS INT
AS
BEGIN

DECLARE @RET INT = 0
IF EXISTS(
SELECT *
FROM tblDSTYPE DS
    JOIN tblCASE C ON DS.DstypeID = C.DstypeID
    JOIN tblCOUNTRY CO ON C.CountryID = CO.CountryID
    JOIN tblAGE AGE ON C.AgeID = AGE.AgeID
WHERE DS.DstypeName = 'Chewing tobacco'
    AND CO.CountryName = 'China'
    AND AGE.AgeGroupName = '15 to 19'
    AND C.CaseValue = 0
)
SET @RET = 1

RETURN @RET
END
GO

ALTER TABLE tblCASE WITH NOCHECK
ADD CONSTRAINT AgeGroup_Tobacco
CHECK (dbo.AgeGroup_Tobacco_Rule() = 0)
GO
```

Views

Determine the country that has a obesity rate **greater than 0.04** in the region of **Asia** in **2014** that also has **at least 0.05** chewing tobacco rate in that year

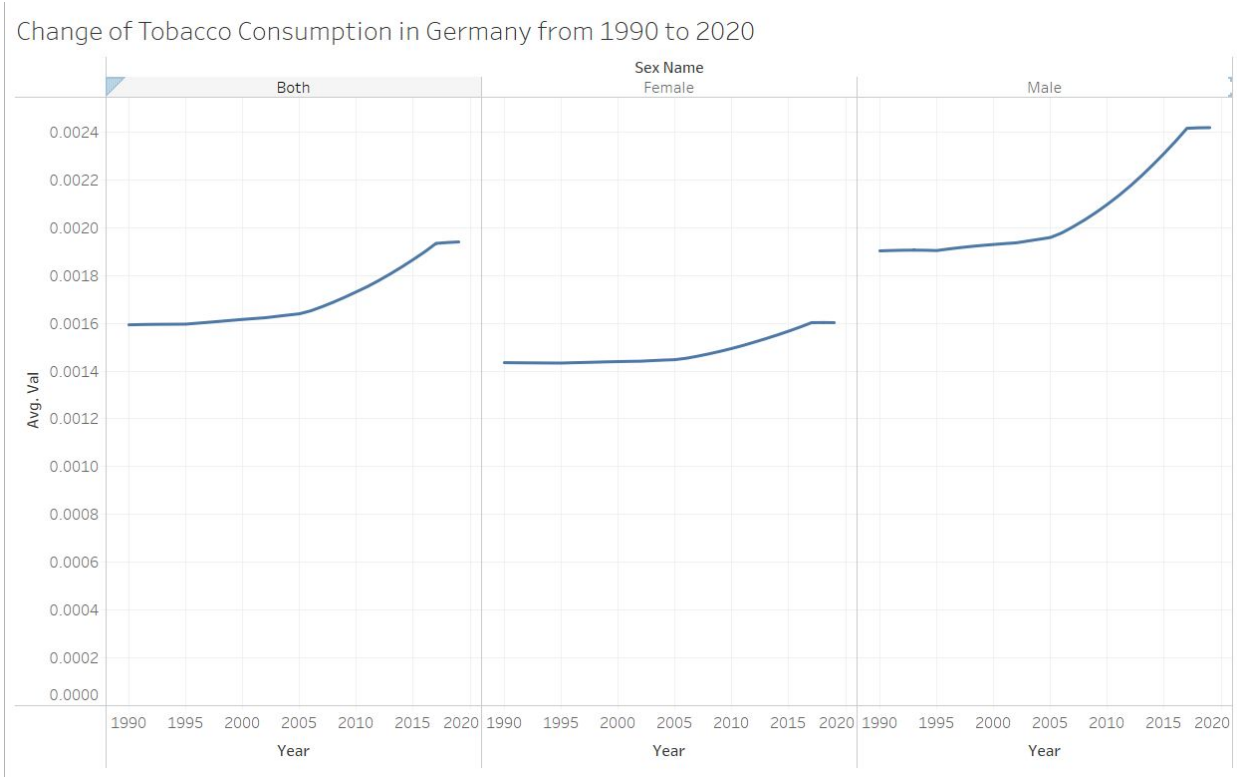
```
SELECT *
FROM
(SELECT CO.CountryID, CO.CountryName
FROM tblCOUNTRY CO
    JOIN tblREGION R ON CO.RegionID = R.RegionID
    JOIN tblCASE C ON CO.CountryID = C.CountryID
    JOIN tblDSTYPE DS ON C.DstypeID = DS.DstypeID
WHERE R.RegionName = 'Asia'
    AND DS.DstypeName = 'Obese'
    AND C.CaseValue > 0.04
    AND C.CaseYear = '2014'
GROUP BY CO.CountryID, CO.CountryName) A,

(SELECT CO.CountryID, CO.CountryName
FROM tblCOUNTRY CO
    JOIN tblCASE C ON CO.CountryID = C.CountryID
    JOIN tblDSTYPE DS ON C.DstypeID = DS.DstypeID
WHERE DS.DstypeName = 'Chewing tobacco'
    AND C.CaseYear = '2014'
    AND C.CaseValue >= 0.05
GROUP BY CO.CountryID, CO.CountryName) B

WHERE A.CountryID = B.CountryID
GO
```

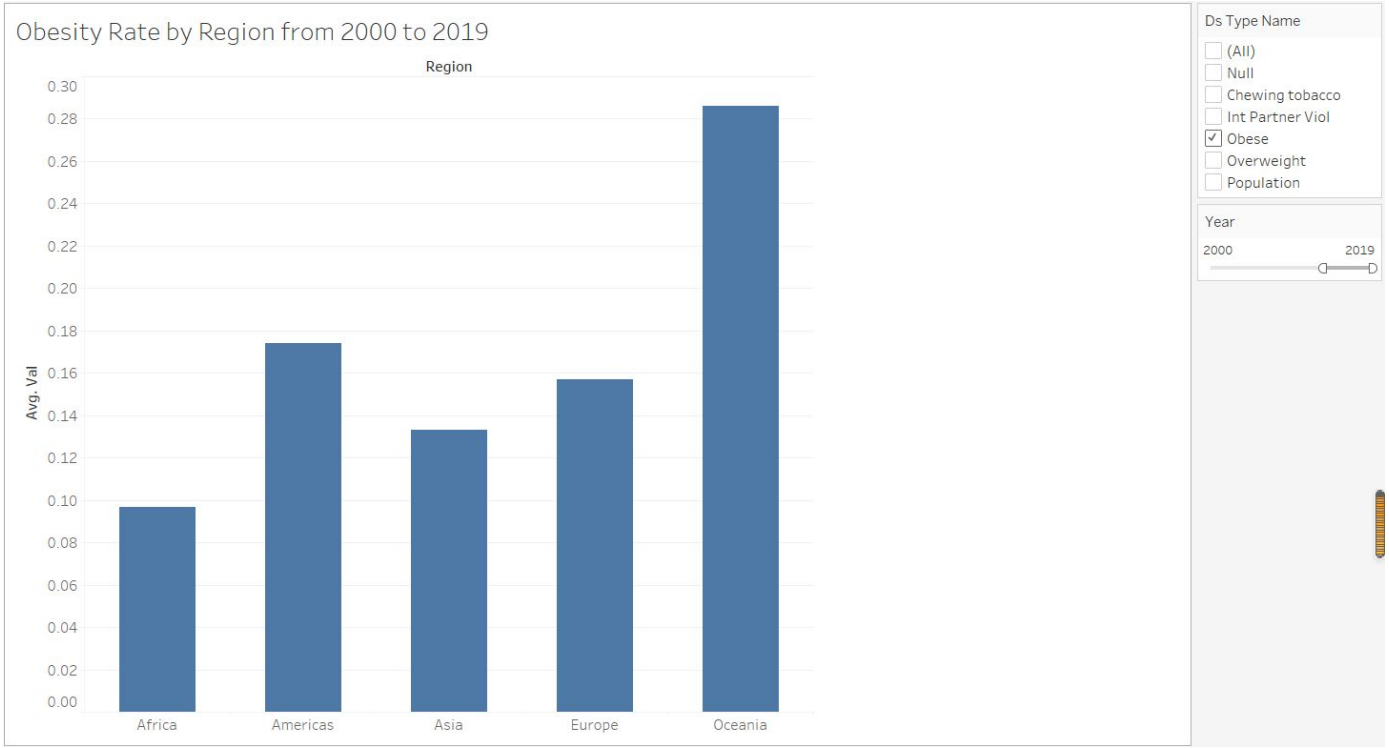
Data Visualization

Grouped Line Chart



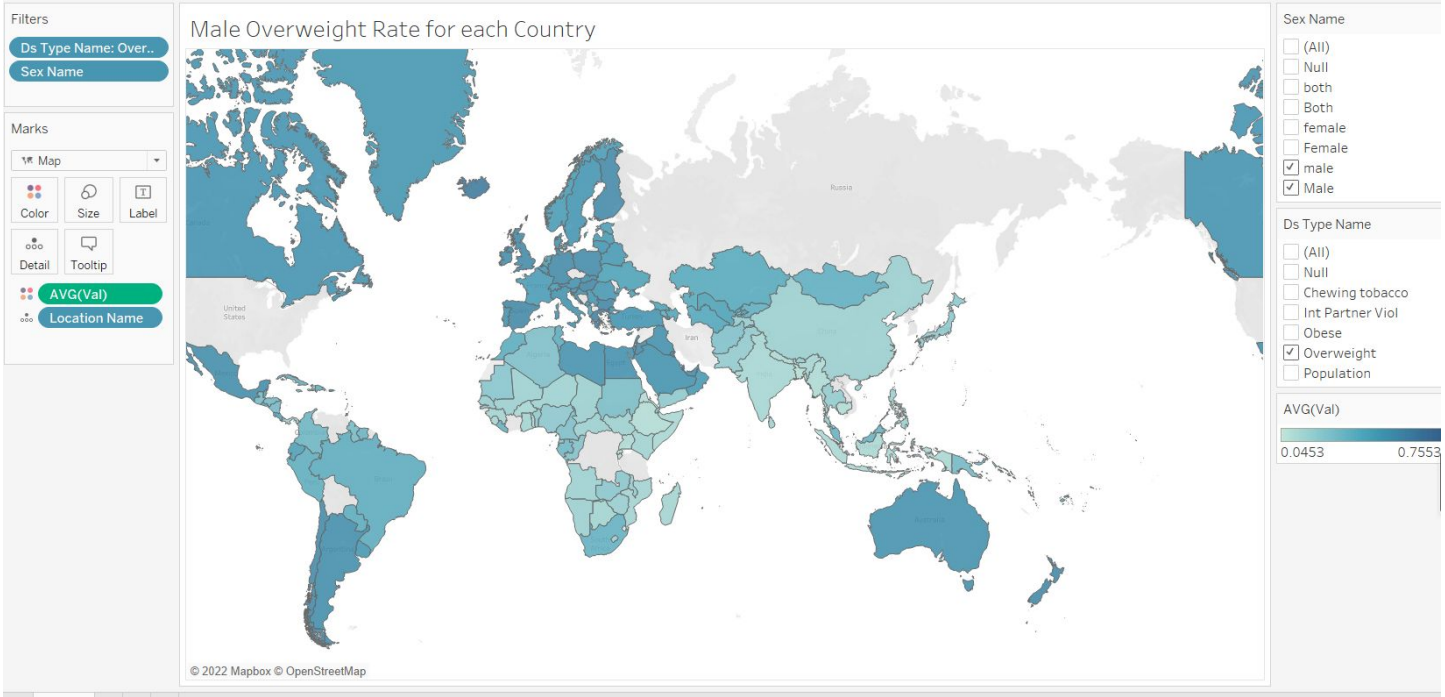
Data Visualization

Bar Chart



Data Visualization

Map



Thank You