```
/* Part1 */

/* 1. Remove years _1995 through _2013. */

/* 2. Create the country_name and tourism_type columns */

/* Part2 */

/* 3. convert values to uppercase and convert '..' to missing values */

/* 4. determine the conversion type */

/* 5. change the data not available in _2014 to a single ".".*/

/* Part3 */

/* 6. Create the Y2014 column and change the original values in_2014 and multiplying by the conversion type*/

/* 7. create the new category column and change the original values to the required values */

/* 8. Permenantly format Y2014*/

/* 9. Remove unnecessary variables*/



%let path = /home/u41140628/EPG194/ECRB94/data;

libname cr "&path/output";


data cleaned_tourism;

        length country_name $300 tourism_type $20;

        retain country_name "" tourism_type "";

        set cr.Tourism(drop=_1995-_2013);

        if A ne . then country_name=country;

        if lowcase(country)="inbound tourism" then tourism_type="Inbound tourism";

                else if lowcase(country)='outbound tourism' then tourism_type="Outbound tourism";

        if country_name ne country and country ne tourism_type;

        series=upcase(series);

        if series = ".." then series="";

        conversion_type=scan(country,-1,"");
```

```
        if _2014=".." then _2014=".";
        if conversion_type="Mn" then do;
                if _2014 ne "." then Y2014=input(_2014,16.)*1000000;
                        else Y2014=.;
                category=cat(scan(country,1,'-','r'),' -US$');
        end;
        else if conversion_type="Thousands" then do;
                if _2014 ne "." then Y2014=input(_2014,16.)*1000000;
                        else Y2014=.;
                category=scan(country,1,'-','r');
        end;
        drop A conversion_type country _2014
run;

proc freq data =cleaned_tourism;
        tables country_name tourism_type series conversion_type;
run;

proc freq data =cleaned_tourism;
        tables country category;
run;

proc freq data =cleaned_tourism;
        tables tourism_type series category;
run;

proc means data =cleaned_tourism mean min max n maxdec=0;
        var Y2014;
run;
```

```
/* Create custom Format */

proc format;
	values contID
	1="North America"
	2="South America"
	3="Europe"
	4="Africa"
	5="Asia"
	6="Oceania"
	7="Antartica";

/* Mere Matching Rows */

proc sort data=cr.country_info(rename=(country=country_name)) out=country_sorted;
	by country_name;
run;

data final_tourism;
	merge cleaned_tourism(in=t) country_sorted(in=c);
	by country_name;
	if t=1 and c=1 then output final_tourism;
	format continent contID.;
run;

proc freq data =final_tourism nlevels;
	tables category tourism_type series continent / nocum nopercent;
run;
```

```
proc means data =final_tourism min mean max n maxdec=0;
        var Y2014;
run;

/* Create the nocountry found table*/

data final_tourism NoCountryFound(keep=country_name);
        merge cleaned_tourism(in=t) country_sorted(in=c);
        by country_name;
        if t=1 and c=1 then output final_tourism;
        if (t=1 and c=0) and first.country_name=1 then output NoCountryFound;
        format continent contID.;
run;

proc means data=final_tourism mean min max n maxdec=0;
        var y2014;
        class Continent;
        where Category="Arrivals";
run;

proc means data=final_tourism mean maxdec=0;
        var y2014;
        where lowcase(Category) contains "tourism expenditure in other countries";
run;
```