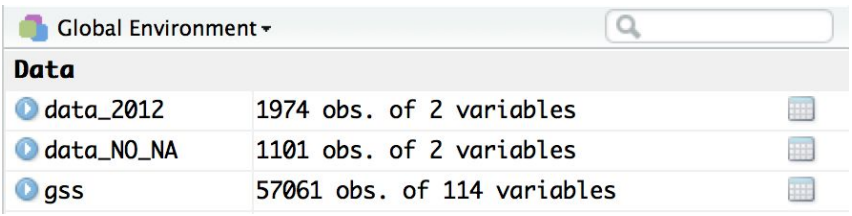
Exploratory analysis

After filtering the data for the year 2012 and removing the rows in which either of the research variables *homosex* or *income06* had an 'NA' value, a data set with 1101 observations was obtained.

```
data_NO_NA <- subset(gss, gss$year == 2012 & gss$homosex != "NA" & gss$income06 != "NA" & gss$income != “Refused”, select = c(homosex, income06))
```




```
> head(data_NO_NA)
                homosex           income06
55088 Not Wrong At All    $150000 Or Over
55089     Always Wrong    $150000 Or Over
55090 Not Wrong At All $110000 To $129999
55093 Not Wrong At All    $25000 To 29999
55095 Not Wrong At All    $5 000 To 5 999
55096 Not Wrong At All       Under $1 000
```

From that on, the following descriptive statistics were computed:

For homosex:

- Counts

Value	Count
Always Wrong	485
Almst Always Wrg	35
Sometimes Wrong	73
Not Wrong At All	508
Other	0

For income06:

- Counts

Value	Count		Value	Count
Under $1 000	19		$22500 To 24999	46

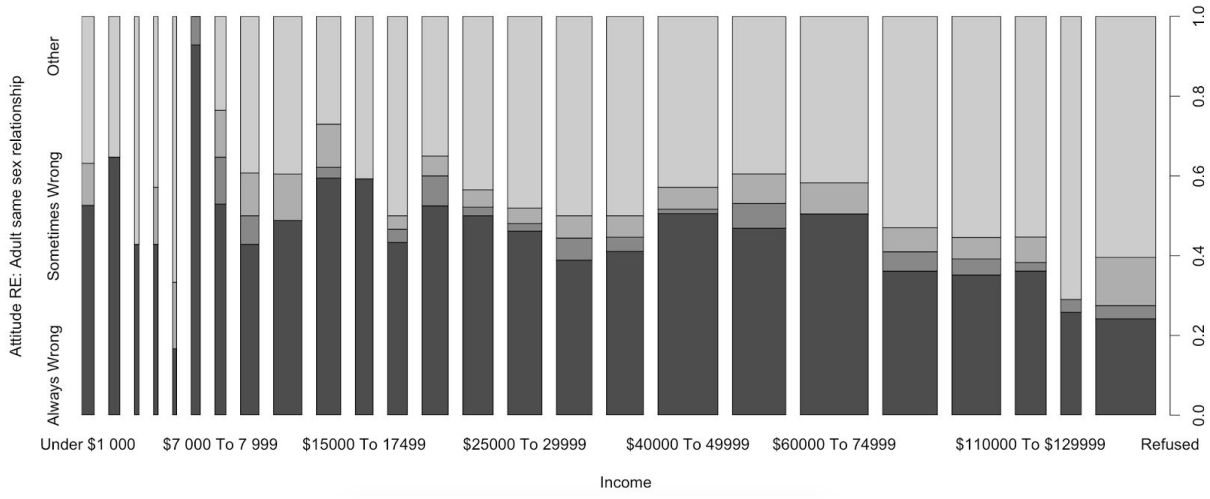
$1 000 To 2 999	17		$25000 To 29999	52
$3 000 To 3 999	7		$30000 To 34999	54
$4 000 To 4 999	7		$35000 To 39999	56
$5 000 To 5 999	6		$40000 To 49999	91
$6 000 To 6 999	14		$50000 To 59999	81
$7 000 To 7 999	17		$60000 To 74999	103
$8 000 To 9 999	28		$75000 To $89999	83
$10000 To 12499	43		$90000 To $109999	74
$12500 To 14999	37		$110000 To $129999	47
$15000 To 17499	27		$130000 To $149999	31
$17500 To 10999	30		$150000 Or Over	91
$20000 To 22499	40		Refused	0

Combining the two variables

The counts of the separated variables do not provide meaningful information for the analysis. Combining the two variables resulted in the following table:

	Always Wrong	Almst Always Wrg	Sometimes Wrong	Not Wrong At All	Other
Under $1 000	10	0	2	7	0
$1 000 To 2 999	11	0	0	6	0
$3 000 To 3 999	3	0	0	4	0
$4 000 To 4 999	3	0	1	3	0
$5 000 To 5 999	1	0	1	4	0
$6 000 To 6 999	13	1	0	0	0
$7 000 To 7 999	9	2	2	4	0
$8 000 To 9 999	12	2	3	11	0
$10000 To 12499	21	0	5	17	0
$12500 To 14999	22	1	4	10	0
$15000 To 17499	16	0	0	11	0
$17500 To 19999	13	1	1	15	0
$20000 To 22499	21	3	2	14	0
$22500 To 24999	23	1	2	20	0
$25000 To 29999	24	1	2	25	0
$30000 To 34999	21	3	3	27	0
$35000 To 39999	23	2	3	28	0
$40000 To 49999	46	1	5	39	0
$50000 To 59999	38	5	6	32	0
$60000 To 74999	52	0	8	43	0
$75000 To $89999	30	4	5	44	0
$90000 To $109999	26	3	4	41	0
$110000 To $129999	17	1	3	26	0
$130000 To $149999	8	1	0	22	0
$150000 Or Over	22	3	11	55	0
Refused	0	0	0	0	0

Which produced this plot:



The distribution does not seem to show a clear correlation between the two variables. There are some ranges in which one of the two extreme answers - "*Always wrong"* and "*Not wrong at all"* regarding the position of the participant towards same sex relationships - is stronger than the other, but there doesn't seem to be a direct correlation, positive or negative, between levels of income and attitude towards homosexual relationships.