

Anova Test Assignment

Anova Test

```
library(ggplot2)
```

```
file.path <- "/cloud/project/Data/Dataset.csv"  
Dataset.csv <- read.csv(file.path)  
head(Dataset.csv)
```

	S.NoSort.column	Case.Identification.Number	CONTROL..COMPLETES.1.PHONE.2....
1	NA	CASEIDSort column	MODESort column
2	1	1	(2) Mail
3	2	2	(9) Web
4	3	3	(9) Web
5	4	4	(9) Web
6	5	5	(1) Phone

	CENSUS.STATE.FIPS.CODES.LABE....	Added..Census.Region	Added..Census.Division
1	STFIPSSort column	REGIONSort column	DIVISIONSort column
2	(51) VIRGINIA	(3) South	(5) South Atlantic
3	(6) CALIFORNIA	(4) West	(9) Pacific
4	(28) MISSISSIPPI	(3) South	(6) East South Central
5	(36) NEW YORK	(1) Northeast	(2) Mid-Atlantic
6	(18) INDIANA	(2) Midwest	(3) East North Central

	OFFERED.INTAKE.ASS.MNT.REFER....	Offered.Detox.3.31.03	Offered.SA.Tx.3.31.03
1	OTHNONTXSort column	DETOXSort column	TREATMTSort column
2	(1) Yes	(0) No	(1) Yes
3	(1) Yes	(0) No	(1) Yes
4	(1) Yes	(1) Yes	(1) Yes
5	(0) No	(0) No	(1) Yes
6	(1) Yes	(0) No	(1) Yes

	Setting..Halfway.House	OFFERED.OTHER.SA.SERVICES.SU....
--	------------------------	----------------------------------

1	LOC5Sort column	ADMINSort column
2	(1) Yes	(1) Yes
3	(1) Yes	(1) Yes
4	(0) No	(0) No
5	(0) No	(0) No
6	(0) No	(0) No

PRIMARY.FOCUS.SA.TX.MH.MIX.G....

1	FOCUSSort column
2	(1) Substance abuse treatment services
3	(1) Substance abuse treatment services
4	(3) Mix of mental health and substance abuse
5	(1) Substance abuse treatment services
6	(3) Mix of mental health and substance abuse

Ownership Federal.Government.Agency

1	OWNERSHPSort column	FEDOWNSort column
2	(2) Private non-profit organization	<NA>
3	(2) Private non-profit organization	<NA>
4	(1) Private for-profit organization	<NA>
5	(4) Local, county, or community government	<NA>
6	(2) Private non-profit organization	<NA>

Solo.practice AFFILIATED.W..RELIGIOUS.ORG....

1	LOC15Sort column	RELIGSort column
2	(0) No	(0) No
3	(0) No	(0) No
4	(0) No	(0) No
5	<NA>	<NA>
6	(0) No	(0) No

LOCATED.IN.OPERATED.BY.HOSPI....

1	HOSPITALSort column	Hospital.Type Hotline...yes.no
2	(0) No	<NA> (0) No
3	(0) No	<NA> (0) No
4	(1) Yes (2) Psychiatric hospital	(1) Yes
5	(0) No	<NA> (0) No
6	(0) No	<NA> (0) No

Assessment.comprehensive.SA Assessment.mental.health

1	SRVC1Sort column	SRVC2Sort column
2	(0) No	(0) No
3	(1) Yes	<NA>
4	(1) Yes	(1) Yes
5	(1) Yes	<NA>
6	(1) Yes	(1) Yes

Therapy.family.counseling Therapy.group

1	SRVC4Sort column	SRVC5Sort column
---	------------------	------------------

2	(0) No	(1) Yes
3	(1) Yes	(1) Yes
4	(1) Yes	(1) Yes
5	<NA>	(1) Yes
6	(1) Yes	(1) Yes

```
file_path <- "/cloud/project/Data/Dataset.csv"
df <- read.csv(file_path, header = TRUE)
```

We are using Therapy.group as a dependent variable and Added..Census.Region as an independent variable

```
file_path <- "/cloud/project/Data/Dataset.csv"
df <- read.csv(file_path, header = TRUE)
Dataset.csv <- Dataset.csv[-2, ]
# Remove duplicate header
```

convert "Therapy.group" to numeric (1=yes and 0= No)

```
Dataset.csv$therapy_group_num <- ifelse(Dataset.csv$Therapy.group == "(1) Yes", 1,
                                         ifelse(Dataset.csv$Therapy.group == "(0) No", 0, NA))
```

```
Dataset.csv_clean <- subset(Dataset.csv,
                           !is.na(therapy_group_num) &
                           Added..Census.Region != "REGIONSort column" &
                           !is.na(Added..Census.Region))

Dataset.csv_clean$region <- Dataset.csv_clean$Added..Census.Region

table(Dataset.csv_clean$region)
```

(0) US Jurisdiction/Territory	(1) Northeast
1	9
(2) Midwest	(3) South
3	5
(4) West	
6	

```
anova_model <- aov(therapy_group_num ~ region, data = Dataset.csv_clean)
summary(anova_model)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
region	4	0.5333	0.1333	0.905	0.481
Residuals	19	2.8000	0.1474		

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

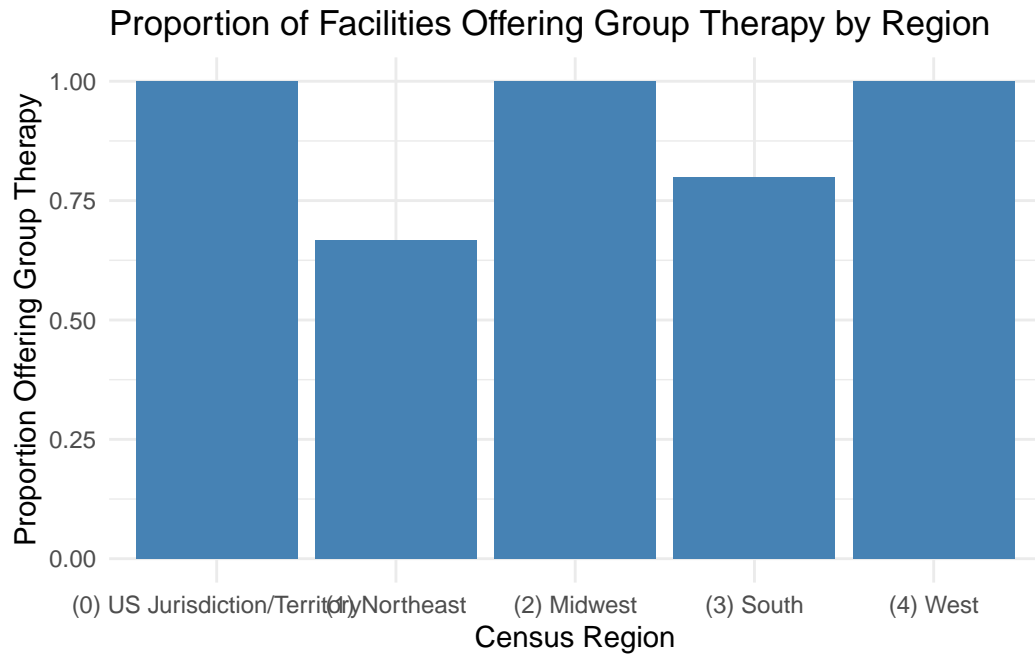
filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
plot_data <- Dataset.csv_clean %>%
  group_by(region) %>%
  summarize(
    n = n(),
    therapy_yes = sum(therapy_group_num == 1, na.rm = TRUE),
    proportion = therapy_yes / n
  )

ggplot(plot_data, aes(x = region, y = proportion)) +
  geom_bar(stat = "identity", fill = "steelblue") +
  labs(title = "Proportion of Facilities Offering Group Therapy by Region",
       x = "Census Region",
       y = "Proportion Offering Group Therapy") +
  theme_minimal() +
  ylim(0, 1)
```



A one-way ANOVA was conducted to compare the effect of census Region on whether therapy group services were offered. There was no significant difference in therapy group availability across the five regions, The F-value is 0.905, and the p-value is 0.481, because of $p > 0.05$ the result is not significant. This means that there is no evidence that the proportion of facilities offering therapy groups services differs by census region in this data set.