R Lab2:A Brief Introduction to R(continued)

Haocheng Hu and Zhaopeng Qu

Nanjing University

3/5/2025

1 Data Practice

2 Plot

Section 1

Data Practice

Importing Data: From CSV & Excel

```r
caschool_csv <- read.csv("Data/caschool.csv")
```

Importing Data: From Other Statistical Tools

Funtion	Format
read.spss	SPSS
read.dta	Stata
read.ssd	SAS
read.mtp	minitab

Importing Data: From STATA

```
library("foreign")
caschool <- read.dta("Data/caschool.dta")
```

View Data:

```
#View(caschool)
head(caschool)
```

Drop Variables

```
ca_data_small <- subset(caschool,select=c(observat,testscr,+
                                          str,expn_stu,el_pct,avginc))
```

Generate New Variables

```
ca_data_small$logexp <- log(ca_data_small$expn_stu)
ca_data_small$el_high <- ca_data_small$el_pct
head(ca_data_small)
```

```
##   observat testscr      str expn_stu    el_pct     avginc
## 1        1  690.80 17.88991 6384.911  0.000000 22.690001 8.
## 2        2  661.20 21.52466 5099.381  4.583333  9.824000 8.
## 3        3  643.60 18.69723 5501.955 30.000002  8.978000 8.
## 4        4  647.70 17.35714 7101.831  0.000000  8.978000 8.
## 5        5  640.85 18.67133 5235.988 13.857677  9.080333 8.
## 6        6  605.55 21.40625 5580.147 12.408759 10.415000 8.
```

Drop Variables

```
ca_data_small <- subset(ca_data_small,select=-c(el_high))
head(ca_data_small)
```

```
##   observat testscr      str expn_stu    el_pct     avginc
## 1        1  690.80 17.88991 6384.911  0.000000 22.690001 8.
## 2        2  661.20 21.52466 5099.381  4.583333  9.824000 8.
## 3        3  643.60 18.69723 5501.955 30.000002  8.978000 8.
## 4        4  647.70 17.35714 7101.831  0.000000  8.978000 8.
## 5        5  640.85 18.67133 5235.988 13.857677  9.080333 8.
## 6        6  605.55 21.40625 5580.147 12.408759 10.415000 8.
```

Summary the Data

```
summary(ca_data_small)
```

```
##     observat        testscr           str           expn_st
##  Min.   :  1.0   Min.   :605.5   Min.   :14.00   Min.   :39
##  1st Qu.:105.8   1st Qu.:640.0   1st Qu.:18.58   1st Qu.:49
##  Median :210.5   Median :654.5   Median :19.72   Median :52
##  Mean   :210.5   Mean   :654.2   Mean   :19.64   Mean   :53
##  3rd Qu.:315.2   3rd Qu.:666.7   3rd Qu.:20.87   3rd Qu.:56
##  Max.   :420.0   Max.   :706.8   Max.   :25.80   Max.   :77
##      el_pct          avginc          logexp
##  Min.   : 0.000   Min.   : 5.335   Min.   :8.275
##  1st Qu.: 1.941   1st Qu.:10.639   1st Qu.:8.498
##  Median : 8.778   Median :13.728   Median :8.559
##  Mean   :15.768   Mean   :15.317   Mean   :8.571
##  3rd Qu.:22.970   3rd Qu.:17.629   3rd Qu.:8.631
##  Max.   :85.540   Max.   :55.328   Max.   :8.950
```

Summary a Variable

```
summary(ca_data_small$testscr)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   605.5   640.0   654.5   654.2   666.7   706.8
```

Attach a Dataframe

- if the dataframe is attached, simply

```
attach(ca_data_small)
summary(testscr)
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    605.5   640.0   654.5   654.2   666.7   706.8
```

```
detach(ca_data_small)
```

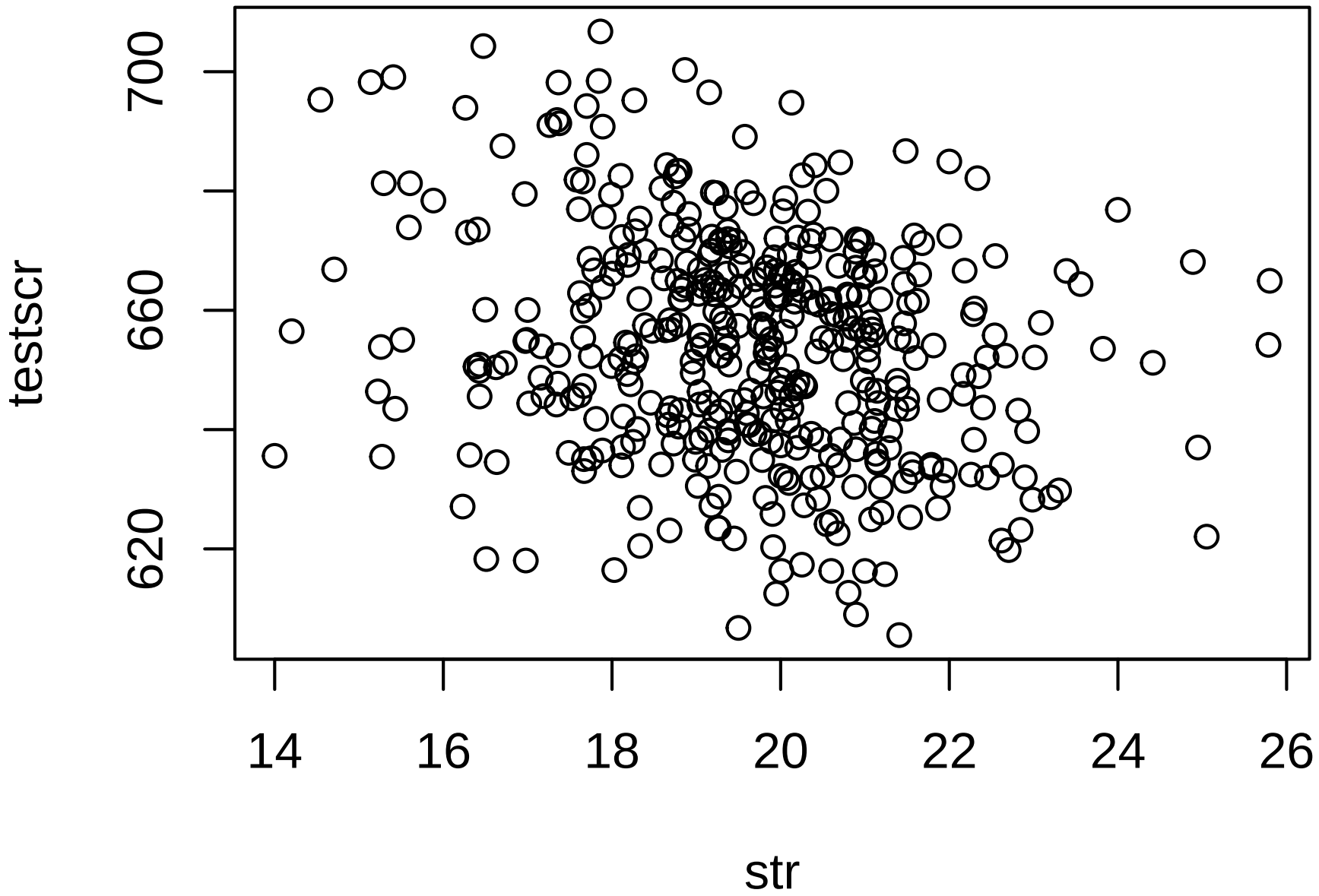
Section 2

Plot

Scatter Plot

```
attach(ca_data_small)
plot(str, testscr)
```

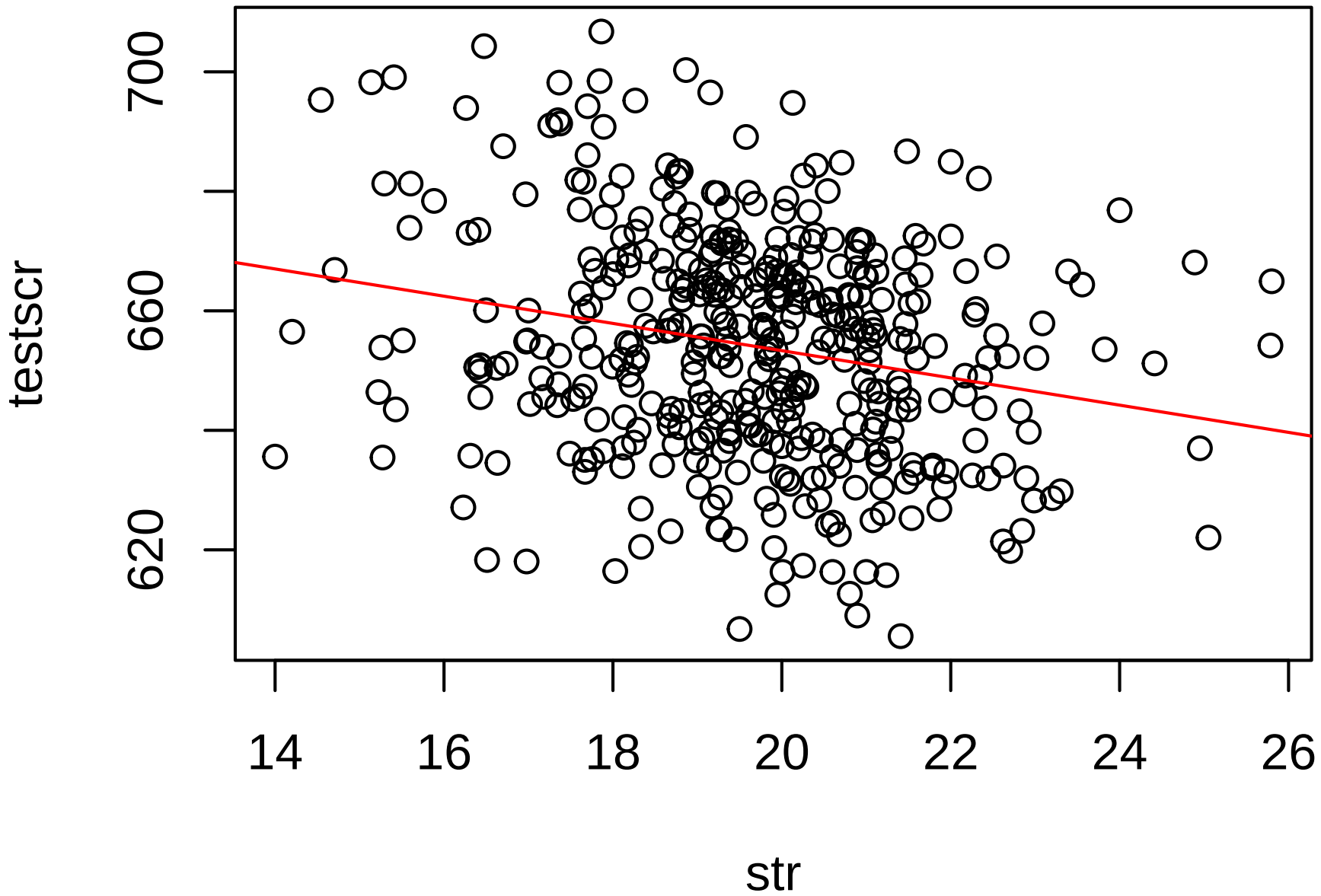
Scatter Plot



Scatter Plot

```
plot(str, testscr)
abline(lm(testscr ~ str , data = ca_data_small),col = "red")
```

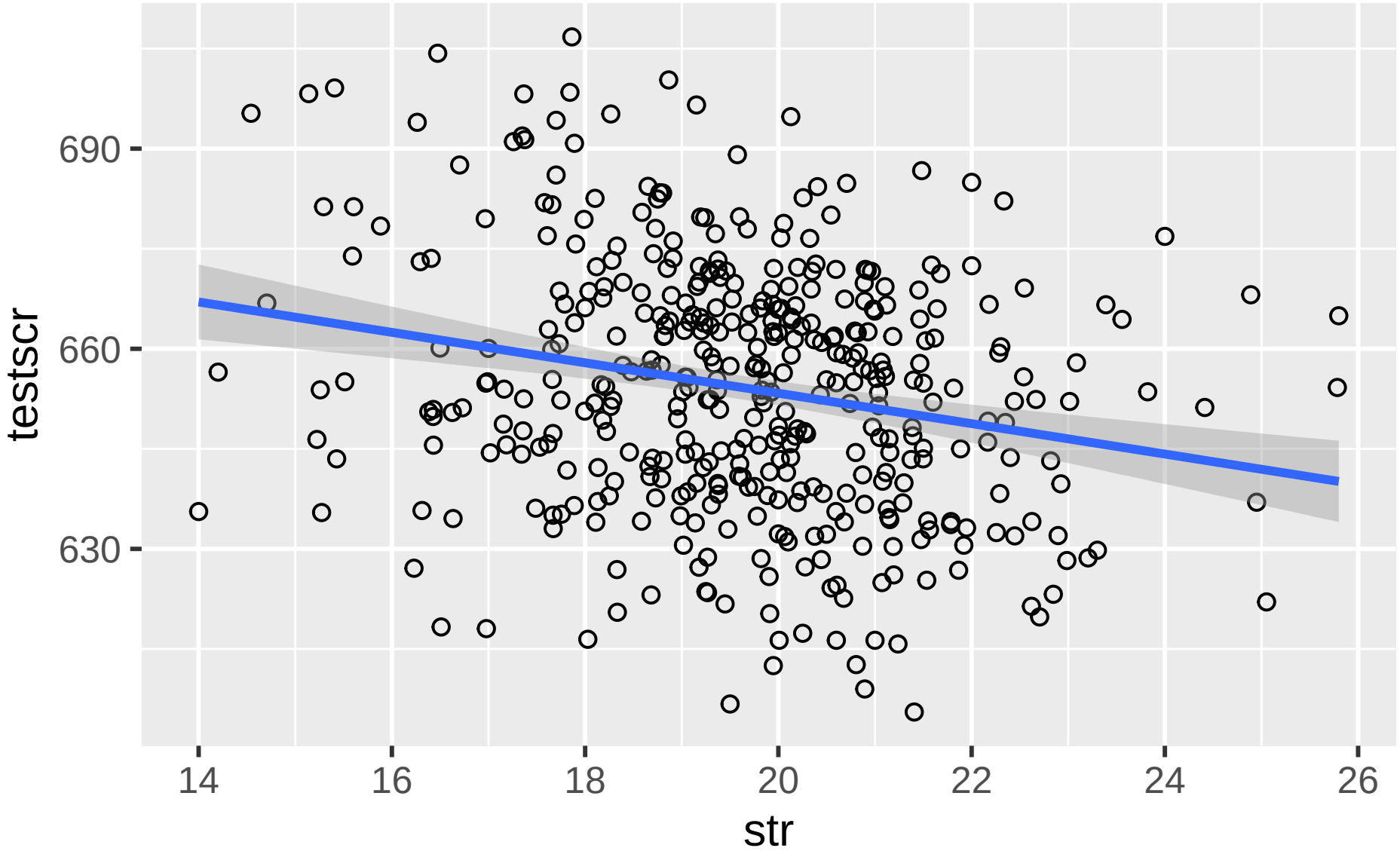
ggplot2



ggplot2

```
library("ggplot2")
ggplot(data =ca_data_small,aes(x=str, y=testscr)) +
geom_point(shape=1) + # Use hollow circles
geom_smooth(method=lm) # Add linear regression line
```

ggplot2



Reference

- Jared P. Lander (2013), *R for Everyone: Advanced Analytics and Graphics*
- Robert I. Kabacoff(2011), *R in Action: Data Analysis and Graphics With R*
- 谢益辉等 (2018),*R* 语言忍者秘籍