



 $0.7160656 - 0.3797124 \ 0.7115967 - 0.1889281 - 0.7032382 - 1.1096595 - 0.3363823 \ 0.8176170 \ 0.1913517 - 0.2098376 \ [251]$ 08228 -1.2692307 1.2485665 -1.0977928 0.6821302 -0.6476787 1.7315770 -1.6743805 0.9240398 0.1716021 [261] 0.851 71520 0.8063297 -0.1635424 0.4145850 0.6619074 -0.4082618 1.0023536 0.2962587 -0.0455598 [271] -0.3377601 0.009 29061 0.0804276 -1.1327947 -0.4724080 -0.5483509 1.4254622 1.0142226 -0.7548832 [281] -0.4243025 0.2201248 0.027 75315 -1.4021574 -1.7079080 1.2708254 0.5252110 0.0944564 -0.1857770 [291] 0.8227716 -0.0343559 0.7184933 -0.550 85377 0.3668166 0.7319903 -0.7136734 0.7931802 -1.5419692 [301] -0.3275177 -1.6955021 -0.7664577 2.3103283 -0.537 10885 0.6676794 0.8812027 -0.8826746 -0.0169117 [311] 0.6240759 -0.0820713 0.1483842 -0.1427901 -0.4799147 -0.458 58435 0.2093504 0.0814105 1.8690609 [321] 0.4352119 1.9909011 -0.0255430 -0.7666637 -1.4900247 -1.7458375 -1.035 68887 -0.6450265 -0.6060436 [331] -0.8159328 -0.2177304 0.3557860 -2.3815800 0.4280608 -2.2567250 -1.7111780 1.732 78034 1.3770028 [341] 0.2271645 0.0190696 -0.6397869 0.4446839 1.2774298 -0.3691963 -0.8395816 -0.3240304 0.110 88310 [351] 0.7385059 -0.8239616 -0.6010957 0.6806157 0.0063708 1.0484444 -0.5214221 -0.8688074 -0.6692890 -0.552  $-0.6588511 - 0.8981495 - 1.0673545 \quad 0.7994342 - 0.0669976 - 1.3183245 \quad 0.3702838 - 0.4788287 \quad 1.6122158 \quad 0.9691895 \quad [371] - 0.06588511 - 0.8981495 - 1.0673545 \quad 0.7994342 - 0.0669976 - 1.3183245 \quad 0.3702838 - 0.4788287 \quad 1.6122158 \quad 0.9691895 \quad [371] - 0.0673545 \quad 0.7994342 - 0.0669976 - 1.3183245 \quad 0.3702838 - 0.4788287 \quad 1.6122158 \quad 0.9691895 \quad [371] - 0.0673545 \quad 0.7994342 - 0.0669976 - 1.3183245 \quad 0.3702838 - 0.4788287 \quad 1.6122158 \quad 0.9691895 \quad [371] - 0.0673545 \quad 0.7994342 - 0.0669976 - 1.3183245 \quad 0.3702838 - 0.4788287 \quad 0.9691895 \quad [371] - 0.0669976 - 0.9691895 \quad [371] - 0.0669976 \quad [371] - 0.0669976$ 05335 0.6018153 -1.0684628 -0.7096711 -1.3780736 0.9156478 1.1259662 -0.9285121 0.1200722 1.5613053 [381] -0.426 59612 -1.1097167 -1.3435353 -0.0180177 0.8386996 -1.0564775 -0.9475907 1.8754802 0.1076621 [391] 1.6070992 1.036 43806 1 03 103 \.3f \(\beta\); 9 0.8464400\(\delta\)0.0340448 2.0559512 -0.5124115 -1.7289937 [401] 0.3550545 -0.46 23223 0.439 88038 -1.2481095 -1.4009616 -1.3086729 [441] 0.0898067 -0.9238770 -0.7754579 -0.3700475 1.1343146 0.6966690 1.538 97596 -0.6098369 -0.8944564 [451] 0.5413484 -0.8647763 -0.9467811 -0.8160911 0.9284492 1.3716441 -0.4872980 -1.136 67635 1.4448767 [461] 1.3493079 -1.4228588 0.4786978 0.3040695 -0.4150574 0.0462064 -0.6397801 1.2555683 1.6790 48545 [471] 1.2078043 -1.3175601 -0.4902107 0.4611098 -2.3246298 -0.9655982 -0.8413548 -0.2264772 -0.1159337 -1.11: -2.1160452 0.9110626 -0.5646935 -0.6506525 0.5676647 -0.2525807 -0.8341889 -1.1063016 -0.8068980 0.7144335 [491] 24222 0.0070713 0.9334361 0.5080611 -0.5291276 0.3860185 0.2551496 -1.5254860 0.5882182 -0.9391098 [501] -1.576 65546 -0.0850655 0.1205662 0.0492091 1.6740339 -1.3920717 1.8455457 0.8182496 1.0671930 [511] -0.0898167 1.0322 82744 -0.2253102 0.6124574 -0.0553955 1.0423951 0.2987654 0.4683266 1.0143614 [521] 0.3182044 1.4629167 2.3547 53721 0.9055108 -1.3635307 0.9651887 1.0087722 -0.0001732 0.0485673 [531] -0.4870944 0.1478238 -0.5577689 -0.819 91071 0.6207478-0.1193080-0.8003797 0.1568611-0.2461361 [541]-1.7861941 0.9169431 0.1155164 1.2118375-0.500 57990 1.3077573 -0.9011344 2.6798482 0.1874725 [551] -0.9764731 0.0633646 -0.2033034 -0.9134958 -1.4379754 0.428 16124 1.1363646 0.8937146 0.2342962 [561] -0.0825818 -0.0530010 0.0315594 -0.0683545 0.8379897 0.1801809 -0.798: 73510 0.0617739 1.0826551 [571] 0.9841809 -1.2388596 0.8687167 -1.5211849 0.2750131 -0.4145926 -0.7230157 -1.456 <u> 02814 -0.5175049 [581] 1.7952190 -0.9848091 0.7949877 0.3953470 -1.2640616 -0.1460294 -0.4098662 0.0798892 -0.42</u>

# What or How to Summarize Your Data

- It depends on your research question(s)
- Do not make blind ventures with your data
- Avoid overcooking
- Use your intelligence and common sense



#### Distribution!

- Distributional information would be be-all and end-all
- Statistics such as mean, standard deviation, and regression coefficients stand for some partial (but important) features of the distribution
- Visualization is always the first step of your analysis









Summarizing Data

## **Summary Statistics**

- Location (of the distribution): mean, median, trimmed mean...
- Scale (of the distribution): standard deviation, variance, range, IQR...
- Shape (of the distribution): skewness, kurtosis...
- Relative location (in the distribution): quantile, percentile, minimum, maximum..



```
library(MatchIt)
data(lalonde)
attach(lalonde)

boxplot(re78~treat)

re78.treat <- re78[treat==1]
re78.contr <- re78[treat==0]

mean(re78); sd(re78)
mean(re78.treat); sd(re78.treat)
mean(re78.contr); sd(re78.contr)

summary(lalonde)</pre>
```

#### Summarizing Data

```
summary.stats <- function(y)</pre>
    x <- na.omit(y)
                         # Omit missing values
    m \leftarrow mean(x)
    s < - sd(x)
    z < - (x-m)/s
                          # Standardization
    skew <- mean(z^3)
                         # Skewness
    kurt <- mean(z^4)
                         # Kurtosis
    mini <- min(x)
                          # Minimum
    maxi <- max(x)</pre>
                          # Maximum
    q \leftarrow quantile(x, probs=c(.25, .50, .75))
    res <- list(average=m, stdev=s,</pre>
                 skewness=skew, kurtosis=kurt,
                  q1=q[1], q2=q[2], q3=q[3],
                  minimum=mini, maximum=maxi)
    return(res)
    summary.stats(re78.treat)
    summary.stats(re78.contr)
Summarizing Data
```

```
table(treat, black)
table(treat, hisp)
table(treat, married)
table(treat, nodegr)
par(mfrow=c(1,2))
boxplot(age~treat, names=c("Control", "Treatment"), ylab="age")
boxplot(educ~treat, names=c("Control", "Treatment"),
ylab="educ")
```

## Summarizing Data