Midterm Solutions

STAT 471/571/701

Nov 1, 2016, 6 - 8:10 PM

Contents

Name: _____

Note

These solutions were done pretty quickly, so please excuse my brevity.

Question 1: The College Dropout

Read in the Newsweek dataset. Here are a few lines to aid you. Notice that Penn is in **row 820**.

```r
# Required libraries - you may add more packages as desired
library(leaps)  # regsubsets() for model selection
library(car)  # Anova()
library(glmnet)  # glmnet() and cv.glmnet()
```

```r
rm(list = ls())  # Remove all the existing variables
college_data <- read.csv("USNEWS_graduation_subset.csv")
college_data$Schooltype <- as.factor(college_data$Schooltype)  # 1 - public, 2- private
str(college_data)
dim(college_data)
summary(college_data)
tail(college_data, 10)
head(college_data, 20)
View(college_data)
# <<<<<<<<<< NA VALUES >>>>>>>>>
sum(is.na(college_data))
# show how many NA values in each column
sapply(college_data, function(x) sum(is.na(x)))
names(college_data)
```

We also identify and isolate Penn within this dataset, to use for prediction purposes later.

```r
penn_loc <- which(college_data$Name == "University of Pennsylvania")  # identify Penn
Penn <- college_data[penn_loc, ]
Penn
```

```
##                             Name State Schooltype All.test.std App.accept
## 820 University of Pennsylvania    PA            2     2.423741       5232
##     Acc.Rate Pct.Yield Total.students Student.Faculty Grad.rate
## 820 42.21397   47.0948           9736             6.3        93
##     Pct.fac.degree In.Tuition Room.board
## 820             96      17020       7270
```

a) How many colleges are included in this dataset? How many variables are there in this data set? List all the variable names. Indicate which variables are categorical variables rather than continuous. Make sure they are treated as factors in `college_data`.

There are 1042 entries with 13 variables (n > p by a decent margin, which is good) in this data set. Due to the cleaning there are no 'na' values present.

How many colleges are included?

```r
college_data %>% select(Name) %>% distinct()  # get distinct values for state column
```

```
##                              Name
## 1         Alaska Pacific University
## 2        Alabama Agri. & Mech. Univ.
## 3              Faulkner University
## 4            University of Montevallo
## 5       Auburn University-Main Campus
## 6         Birmingham-Southern College
## 7                Huntingdon College
## 8         Jacksonville State University
```

```
## 9                               Judson College
## 10                          Livingston University
## 11                           University of Mobile
## 12                             Samford University
## 13                           Spring Hill College
## 14                  Troy State University at Troy
## 15           University of Alabama at Tuscaloosa
## 16           University of Alabama at Birmingham
## 17           University of Alabama at Huntsville
## 18                  University of South Alabama
## 19             Auburn University at Montgomery
## 20              Arkansas College (Lyon College)
## 21                      Arkansas Tech University
## 22                    Arkansas State University
## 23            University of Central Arkansas
## 24                               Hendrix College
## 25                        John Brown University
## 26         University of Arkansas at Little Rock
## 27                          Harding University
## 28                       Grand Canyon University
## 29         Arizona State University Main campus
## 30                 Northern Arizona University
## 31                        University of Arizona
## 32                            Biola University
## 33                    California Baptist College
## 34            California Institute of Technology
## 35              California Lutheran University
## 36          California State Univ. at Fullerton
## 37            California Polytechnic-San Luis
## 38                     California Poly-Pomona
## 39              Humboldt State University
## 40   California State University at Sacramento
## 41               San Diego State University
## 42   California State University at Northridge
## 43            San Francisco State University
## 44                  San Jose State University
## 45                  Sonoma State University
## 46   United States International University
## 47                       Chapman University
## 48              Claremont McKenna College
## 49                     Harvey Mudd College
## 50                            Pitzer College
## 51                          Pomona College
## 52                         Scripps College
## 53                 College of Notre Dame
## 54                     Holy Names College
## 55       Dominican College of San Rafael
## 56                    La Sierra University
## 57                 University of La Verne
## 58                            Mills College
## 59                 Mount St. Marys College
## 60                     Occidental College
## 61              Pacific Christian College
## 62                 Fresno Pacific College
```

```
## 63                       Pacific Union College
## 64                 Point Loma Nazarene College
## 65                       Pepperdine University
## 66                 Southern California College
## 67            St. Marys College of California
## 68                         Stanford University
## 69         University of California at Berkeley
## 70            University of California at Davis
## 71           University of California at Irvine
## 72      University of California at Los Angeles
## 73        University of California at Riverside
## 74        University of California at San Diego
## 75    University of California at Santa Barbara
## 76      University of California at Santa Cruz
## 77                     University of Redlands
## 78               University of San Francisco
## 79                     Santa Clara University
## 80             University of Southern California
## 81                  University of the Pacific
## 82                         Westmont College
## 83                         Whittier College
## 84                       Woodbury University
## 85                     University of Judaism
## 86        California State Univ. at Bakersfield
## 87                     University of San Diego
## 88                   Thomas Aquinas College
## 89               Loyola Marymount University
## 90                     Concordia University
## 91                     Adams State College
## 92                         Colorado College
## 93            University of Northern Colorado
## 94                Colorado State University
## 95                       Mesa State College
## 96               Metropolitan State College
## 97                         Regis University
## 98            University of Southern Colorado
## 99          University of Colorado at Boulder
## 100                     University of Denver
## 101         Western State College of Colorado
## 102             University of Colorado at Denver
## 103                 Albertus Magnus College
## 104      Central Connecticut State University
## 105                     Connecticut College
## 106                     Fairfield University
## 107                 University of New Haven
## 108                     Quinnipiac College
## 109                 Sacred Heart University
## 110     Southern Connecticut State University
## 111                   Saint Joseph College
## 112                         Trinity College
## 113               University of Bridgeport
## 114                   University of Hartford
## 115                       Wesleyan University
## 116      Eastern Connecticut State University
```

```
## 117                      Yale University
## 118       University of Connecticut at Storrs
## 119                   American University
## 120          Catholic University of America
## 121              George Washington University
## 122                 Georgetown University
## 123                    Howard University
## 124                Goldey Beacom College
## 125               University of Delaware
## 126                       Wesley College
## 127                      Barry University
## 128             Bethune Cookman College
## 129                St. Thomas University
## 130        Florida Institute of Technology
## 131     Embry Riddle Aeronautical University
## 132            Florida Atlantic University
## 133                       Eckerd College
## 134             Florida Southern College
## 135             Florida State University
## 136                       Lynn University
## 137          Nova Southeastern University
## 138                     Rollins College
## 139                    Stetson University
## 140                  University of Florida
## 141                    University of Miami
## 142            University of South Florida
## 143                       Webber College
## 144         University of Central Florida
## 145           University of West Florida
## 146                      Flagler College
## 147             Warner Southern College
## 148          Palm Beach Atlantic College
## 149          Florida International University
## 150           University of North Florida
## 151                 Agnes Scott College
## 152             Armstrong State College
## 153              Clark Atlanta University
## 154                     Augusta College
## 155                        Berry College
## 156                    Brenau University
## 157                Brewton-Parker College
## 158                    Columbus College
## 159                     Emory University
## 160        Georgia Institute of Technology
## 161        Southern College of Technology
## 162            Georgia Southern University
## 163           Georgia Southwestern College
## 164              Georgia State University
## 165               Kennesaw State College
## 166                     LaGrange College
## 167                    Mercer University
## 168                    Morehouse College
## 169                 Morris Brown College
## 170                 Oglethorpe University
```

```
## 171                       Paine College
## 172                    Piedmont College
## 173              Savannah State College
## 174                    Shorter College
## 175                    Spelman College
## 176               University of Georgia
## 177           Valdosta State University
## 178                   Wesleyan College
## 179               West Georgia College
## 180                    Georgia College
## 181                   Covenant College
## 182      Savannah Coll. of Art and Design
## 183        University of Hawaii at Manoa
## 184         University of Hawaii at Hilo
## 185                 Briar Cliff College
## 186                 Buena Vista College
## 187                     Central College
## 188                      Clarke College
## 189                         Coe College
## 190                     Cornell College
## 191                       Dordt College
## 192                     Drake University
## 193                  Graceland College
## 194                  Grand View College
## 195                    Grinnell College
## 196               Iowa State University
## 197                       Loras College
## 198                      Luther College
## 199        Teikyo Marycrest University
## 200                Morningside College
## 201          Mount Saint Clare College
## 202              Northwestern College
## 203           Saint Ambrose University
## 204         University of Northern Iowa
## 205              University of Dubuque
## 206               Upper Iowa University
## 207                   Wartburg College
## 208         Teikyo Westmar University
## 209            Boise State University
## 210                 Albertson College
## 211             Idaho State University
## 212         Lewis-Clark State College
## 213         Northwest Nazarene College
## 214                University of Idaho
## 215                  Augustana College
## 216                   Aurora University
## 217                       Barat College
## 218                  Blackburn College
## 219                  Bradley University
## 220                   DePaul University
## 221         Eastern Illinois University
## 222                      Eureka College
## 223                 Greenville College
## 224                   Illinois College
```

```
## 225                    Illinois Institute of Technology
## 226                          Illinois State University
## 227                   Northeastern Illinois University
## 228                          Chicago State University
## 229                     Illinois Wesleyan University
## 230                                  Kendall College
## 231                                     Knox College
## 232                              Lake Forest College
## 233                                 Lewis University
## 234                        Loyola University Chicago
## 235                                MacMurray College
## 236                                McKendree College
## 237                               Millikin University
## 238                                Monmouth College
## 239                       National-Louis University
## 240                           North Central College
## 241                               North Park College
## 242                    Northern Illinois University
## 243                         Northwestern University
## 244                                Principia College
## 245                               Quincy University
## 246                             Roosevelt University
## 247   Southern Illinois University at Edwardsville
## 248                 Illinois Benedictine College
## 249                         Saint Xavier University
## 250                            University of Chicago
## 251               University of Illinois - Urbana
## 252               University of Illinois at Chicago
## 253                      Western Illinois University
## 254                                  Wheaton College
## 255                            Anderson University
## 256                            Ball State University
## 257                                Butler University
## 258                                DePauw University
## 259                                  Earlham College
## 260                          University of Evansville
## 261                               Franklin College
## 262                                   Goshen College
## 263                      Grace College and Seminary
## 264                                 Hanover College
## 265                              Huntington College
## 266                       University of Indianapolis
## 267                 University of Southern Indiana
## 268                 Indiana University at Bloomington
## 269                               Indiana Univ. East
## 270   Indiana Univ.-Purdue Univ. at Indianapolis
## 271                         Indiana Univ. at Kokomo
## 272                             Manchester College
## 273                                   Marian College
## 274                      Indiana Wesleyan University
## 275              Purdue University at West Lafayette
## 276             Rose-Hulman Institute of Technology
## 277                           Saint Francis College
## 278                           Saint Josephs College
```

```
## 279                          Saint Mary-of-the-Woods College
## 280                                Saint Marys College
## 281                                 Taylor University
## 282                              Tri-State University
## 283                          University of Notre Dame
## 284                            Valparaiso University
## 285                                  Wabash College
## 286                          Indiana State University
## 287                                 Baker University
## 288                                Bethany College
## 289                                  Bethel College
## 290                               Friends University
## 291                         Pittsburg State University
## 292                          Emporia State University
## 293                          Kansas Wesleyan University
## 294                                McPherson College
## 295                             Southwestern College
## 296                                Sterling College
## 297                                   Tabor College
## 298                              University of Kansas
## 299                          Wichita State University
## 300                         MidAmerica Nazarene College
## 301                              Benedictine College
## 302                               Bellarmine College
## 303                              Spalding University
## 304                                  Centre College
## 305                               Cumberland College
## 306                         Eastern Kentucky University
## 307                               Georgetown College
## 308                          Kentucky State University
## 309                          Kentucky Wesleyan College
## 310                          Morehead State University
## 311                            Murray State University
## 312                            Transylvania University
## 313                                   Union College
## 314                            University of Kentucky
## 315                           University of Louisville
## 316                               Thomas More College
## 317                      Centenary College of Louisiana
## 318                                Dillard University
## 319                           Nicholls State University
## 320                           Louisiana Tech University
## 321            Louisiana State University at Baton Rouge
## 322                           University of New Orleans
## 323                                Loyola University
## 324                            McNeese State University
## 325                        Northeast Louisiana University
## 326            Northwestern State University of Louisiana
## 327                     Southeastern Louisiana University
## 328                     Southern University at New Orleans
## 329                                 Tulane University
## 330                 University of Southwestern Louisiana
## 331                        Xavier University of Louisiana
## 332                   Southern University and A & M College
```

```
## 333                          American International College
## 334                                      Amherst College
## 335                                  Anna Maria College
## 336                                  Assumption College
## 337                                   Merrimack College
## 338                                      Babson College
## 339                                     Bentley College
## 340                                      Boston College
## 341                                   Boston University
## 342                                 Brandeis University
## 343                                    Clark University
## 344                                        Elms College
## 345                          College of the Holy Cross
## 346                                       Curry College
## 347                            Eastern Nazarene College
## 348                                     Emerson College
## 349                                      Gordon College
## 350                                  Harvard University
## 351                                      Lesley College
## 352              University of Massachusetts at Lowell
## 353              Massachusetts Institute of Technology
## 354                           Bridgewater State College
## 355                           North Adams State College
## 356                                  Salem State College
## 357                              Westfield State College
## 358                              Worcester State College
## 359                                Mount Holyoke College
## 360                             Northeastern University
## 361                                  Pine Manor College
## 362                                       Regis College
## 363                                     Simmons College
## 364                                       Smith College
## 365           University of Massachusetts at Dartmouth
## 366                                 Springfield College
## 367                                   Stonehill College
## 368                                   Suffolk University
## 369                                     Tufts University
## 370             University of Massachusetts at Amherst
## 371              University of Massachusetts at Boston
## 372                                   Wellesley College
## 373                         Western New England College
## 374                                    Wheelock College
## 375                                   Williams College
## 376                     Worcester Polytechnic Institute
## 377                                   Hampshire College
## 378                          Simons Rock College of Bard
## 379                  Wentworth Institute of Technology
## 380                                      Capitol College
## 381                             Bowie State University
## 382                 College of Notre Dame of Maryland
## 383                              Columbia Union College
## 384                          Frostburg State University
## 385                                     Goucher College
## 386                                         Hood College
```

```
## 387                          Johns Hopkins University
## 388                                  Loyola College
## 389                         Morgan State University
## 390                       Mount Saint Marys College
## 391                       Salisbury State University
## 392                                St. Johns College
## 393                    St. Marys College of Maryland
## 394                          Towson State University
## 395              University of Maryland at College Park
## 396           University of Maryland at Baltimore County
## 397              University of Maryland Eastern Shore
## 398                               Villa Julie College
## 399                               Washington College
## 400                         Western Maryland College
## 401              University of Maine at Presque Isle
## 402                                     Bates College
## 403                                   Bowdoin College
## 404                                     Colby College
## 405                  University of Maine at Fort Kent
## 406                                     Husson College
## 407                         University of New England
## 408                                    Thomas College
## 409                      University of Maine at Orono
## 410                                      Unity College
## 411                                     Adrian College
## 412                                     Albion College
## 413                                       Alma College
## 414                                 Andrews University
## 415                                     Calvin College
## 416                                  Concordia College
## 417                        Eastern Michigan University
## 418                     Grand Valley State University
## 419                                  Hillsdale College
## 420                                       Hope College
## 421                                 Kalamazoo College
## 422            Lawrence Technological University
## 423                          Michigan State University
## 424             Michigan Technological University
## 425             Lake Superior State University
## 426                                  Oakland University
## 427             Saginaw Valley State University
## 428                             Siena Heights College
## 429                               Spring Arbor College
## 430                     University of Detroit Mercy
## 431              University of Michigan at Dearborn
## 432                            Wayne State University
## 433                       Western Michigan University
## 434                          Center for Creative Studies
## 435                               Northwood University
## 436              University of Michigan at Ann Arbor
## 437                                 Augsburg College
## 438                                  Carleton College
## 439                        College of Saint Benedict
## 440                        College of Saint Catherine
```

```
## 441                       College of St. Scholastica
## 442                        University of St. Thomas
## 443                      Concordia College-Moorhead
## 444                      Gustavus Adolphus College
## 445                            Hamline University
## 446                            Macalester College
## 447                       Moorhead State University
## 448                      Southwest State University
## 449                         Saint Johns University
## 450                  Saint Marys College of Minnesota
## 451                            Saint Olaf College
## 452                 University of Minnesota at Duluth
## 453                 University of Minnesota at Morris
## 454                         Winona State University
## 455              University of Minnesota Twin Cities
## 456                                  Avila College
## 457              Central Missouri State University
## 458                              Columbia College
## 459                         Culver-Stockton College
## 460                                  Drury College
## 461                            Lindenwood College
## 462                            Maryville University
## 463                        Missouri Valley College
## 464                 Missouri Western State College
## 465            Northeast Missouri State University
## 466                                  Park College
## 467                            Rockhurst College
## 468                    Southwest Baptist University
## 469            Southwest Missouri State University
## 470                         Saint Louis University
## 471                            Stephens College
## 472             University of Missouri at Columbia
## 473                University of Missouri at Rolla
## 474             University of Missouri at Kansas City
## 475                        Washington University
## 476                            Webster University
## 477                          Westminster College
## 478                        William Jewell College
## 479                        William Woods University
## 480                        Alcorn State University
## 481                        Blue Mountain College
## 482                            Millsaps College
## 483                          Mississippi College
## 484                Mississippi University for Women
## 485                    Mississippi State University
## 486            Mississippi Valley State University
## 487                                  Rust College
## 488             University of Southern Mississippi
## 489                        William Carey College
## 490                              Carroll College
## 491               Montana State University-Billings
## 492           Montana College of Mineral Sci. & Tech.
## 493                      Montana State University
## 494                        Rocky Mountain College
```

```
## 495                           University of Montana
## 496                          Western Montana College
## 497          North Carolina A. & T. State University
## 498               Appalachian State University
## 499       University of North Carolina at Asheville
## 500                              Barton College
## 501                         Barber Scotia College
## 502                        Belmont Abbey College
## 503                             Bennett College
## 504                          Campbell University
## 505                              Catawba College
## 506                             Davidson College
## 507                              Duke University
## 508                      East Carolina University
## 509               Elizabeth City State University
## 510                                 Elon College
## 511              Fayetteville State University
## 512                    Gardner Webb University
## 513                          Greensboro College
## 514                             Guilford College
## 515                       High Point University
## 516                 Johnson C. Smith University
## 517                        Lenoir-Rhyne College
## 518                            Mars Hill College
## 519                             Meredith College
## 520                            Methodist College
## 521                 Montreat-Anderson College
## 522          North Carolina Central University
## 523          North Carolina Wesleyan College
## 524                   Pembroke State University
## 525                             Pfeiffer College
## 526                               Queens College
## 527                                Salem College
## 528          St. Andrews Presbyterian College
## 529                      St. Augustines College
## 530    North Carolina State University at Raleigh
## 531   University of North Carolina at Chapel Hill
## 532    University of North Carolina at Charlotte
## 533   University of North Carolina at Greensboro
## 534                       Wake Forest University
## 535                        Warren Wilson College
## 536                   Western Carolina University
## 537   University of North Carolina at Wilmington
## 538                              Wingate College
## 539               Winston-Salem State University
## 540                  Dickinson State University
## 541                           Jamestown College
## 542                            University of Mary
## 543                    Mayville State University
## 544                    University of North Dakota
## 545                North Dakota State University
## 546                         Chadron State College
## 547                          Creighton University
## 548                                  Dana College
```

```
## 549                        Doane College
## 550                     Hastings College
## 551        University of Nebraska at Kearney
## 552                    Peru State College
## 553        University of Nebraska at Lincoln
## 554                    Wayne State College
## 555                  Colby-Sawyer College
## 556                     Dartmouth College
## 557               Franklin Pierce College
## 558                New Hampshire College
## 559                  Notre Dame College
## 560                        Rivier College
## 561                  Saint Anselm College
## 562          University of New Hampshire
## 563                  Keene State College
## 564                 Daniel Webster College
## 565        Thomas More Coll. of Liberal Arts
## 566                      Caldwell College
## 567                     Centenary College
## 568             College of Saint Elizabeth
## 569                        Drew University
## 570                Georgian Court College
## 571             Rowan College of New Jersey
## 572              Jersey City State College
## 573              Montclair State University
## 574       New Jersey Institute of Technology
## 575               Kean College of New Jersey
## 576   William Paterson College of New Jersey
## 577                  Princeton University
## 578                        Rider University
## 579        Rutgers State University at Newark
## 580                  Seton Hall University
## 581                  Saint Peters College
## 582         Stevens Institute of Technology
## 583                 Trenton State College
## 584                        Upsala College
## 585        Rutgers State University at Camden
## 586               Rutgers at New Brunswick
## 587          Fairleigh Dickinson University
## 588             Ramapo College of New Jersey
## 589           Stockton College of New Jersey
## 590                   Saint Johns College
## 591                   College of Santa Fe
## 592               College of the Southwest
## 593           New Mexico Highlands University
## 594   New Mexico Institute of Mining and Tech.
## 595              New Mexico State University
## 596             Western New Mexico University
## 597          Univ. of New Mexico-Main Campus
## 598             University of Nevada at Reno
## 599                 Sierra Nevada College
## 600                    Adelphi University
## 601                       Dowling College
## 602                       Alfred University
```

```
## 603                          Bard College
## 604                       Canisius College
## 605                 CUNY-Brooklyn College
## 606                   CUNY - City College
## 607                  CUNY-Hunter College
## 608                 CUNY - Queens College
## 609                    Clarkson University
## 610                     Colgate University
## 611          College of Mount St. Vincent
## 612               College of Saint Rose
## 613                  Columbia University
## 614                       Barnard College
## 615                    Cornell University
## 616                     DYouville College
## 617                        Elmira College
## 618                     Fordham University
## 619                     Hamilton College
## 620                     Hartwick College
## 621     Hobart and William Smith Colleges
## 622                   Hofstra University
## 623                    Houghton College
## 624                          Iona College
## 625                       Ithaca College
## 626                        Keuka College
## 627                   Le Moyne College
## 628                   Manhattan College
## 629               Manhattanville College
## 630                      Marist College
## 631           Marymount College Tarrytown
## 632           Marymount Manhattan College
## 633                       Molloy College
## 634           Mount Saint Mary College
## 635       Nazareth College of Rochester
## 636               New York University
## 637               Niagara University
## 638                       Nyack College
## 639                    Pace University
## 640             Polytechnic University
## 641                     Pratt Institute
## 642     Rensselaer Polytechnic Institute
## 643             Roberts Wesleyan College
## 644   Rochester Institute of Technology
## 645               Sarah Lawrence College
## 646                    Skidmore College
## 647                       Siena College
## 648           St. Bonaventure University
## 649             St. John Fisher College
## 650               St. Johns University
## 651               St. Josephs College
## 652             St. Lawrence University
## 653         St. Thomas Aquinas College
## 654                       SUNY at Albany
## 655                   SUNY at Binghamton
## 656                 SUNY at Stony Brook
```

```
## 657                        SUNY College  at Brockport
## 658                         SUNY College at Buffalo
## 659                        SUNY College at Cortland
## 660                        SUNY College at Fredonia
## 661                         SUNY College at Geneseo
## 662                       SUNY College at New Paltz
## 663                         SUNY College at Oneonta
## 664                        SUNY College  at Oswego
## 665                      SUNY College at Plattsburgh
## 666                         SUNY College at Potsdam
## 667                             Syracuse University
## 668            Utica College of Syracuse University
## 669                         University of Rochester
## 670                                  Vassar College
## 671                                  Wagner College
## 672                                   Wells College
## 673                             Yeshiva University
## 674                              CUNY-York College
## 675              Long Island University at Brooklyn
## 676                        SUNY College at Purchase
## 677                           CUNY-Lehman College
## 678                                 SUNY at Buffalo
## 679                     CUNY- Medgar Evers College
## 680                   New School for Social Research
## 681                    CUNY-College of Staten Island
## 682                              Ashland University
## 683                         Baldwin-Wallace College
## 684                                Bluffton College
## 685                   Bowling Green State University
## 686                              Capital University
## 687                 Case Western Reserve University
## 688                              Cedarville College
## 689                      Cleveland State University
## 690                      College of Mount St. Joseph
## 691             Franciscan University of Steubenville
## 692                             College of Wooster
## 693                               Defiance College
## 694                             Denison University
## 695                             Heidelberg College
## 696                                  Hiram College
## 697                         John Carroll University
## 698                                  Kenyon College
## 699                                 Lourdes College
## 700                                  Malone College
## 701                               Marietta College
## 702                            Mount Union College
## 703                              Muskingum College
## 704                                 Oberlin College
## 705                       Ohio Northern University
## 706                                Ohio University
## 707                        Ohio Wesleyan University
## 708                               Otterbein College
## 709                               Tiffin University
## 710                            University of Akron
```

```
## 711                            University of Cincinnati
## 712                            University of Dayton
## 713                            University of Toledo
## 714                               Ursuline College
## 715                               Walsh University
## 716                          Wittenberg University
## 717                              Xavier University
## 718                   Youngstown State University
## 719             Ohio State University at Columbus
## 720            Mount Vernon Nazarene College
## 721                  Miami University at Oxford
## 722                            Antioch University
## 723               Southern Nazarene University
## 724                             Cameron University
## 725         University of Central Oklahoma
## 726                Oklahoma Baptist University
## 727             Oklahoma Christian University
## 728     University of Sci. and Arts of Oklahoma
## 729                  Oklahoma State University
## 730       Oklahoma Panhandle State University
## 731                            Phillips University
## 732          Southeastern Oklahoma State Univ.
## 733                       University of Oklahoma
## 734                              University of Tulsa
## 735                       Oral Roberts University
## 736                      Western Baptist College
## 737                            George Fox College
## 738                      Lewis and Clark College
## 739                               Linfield College
## 740                        Oregon State University
## 741            Oregon Institute of Technology
## 742                              Pacific University
## 743                                    Reed College
## 744                           University of Oregon
## 745                        University of Portland
## 746                      Warner Pacific College
## 747                        Willamette University
## 748                               Albright College
## 749                             Allegheny College
## 750                               Alvernia College
## 751                                  Beaver College
## 752                             Bryn Mawr College
## 753                           Bucknell University
## 754                                 Cabrini College
## 755                   Carnegie Mellon University
## 756                           Cedar Crest College
## 757                               Chatham College
## 758                        Chestnut Hill College
## 759                           College Misericordia
## 760                     Delaware Valley College
## 761                            Dickinson College
## 762                            Drexel University
## 763                          Duquesne University
## 764                                 Eastern College
```

```
## 765                        Elizabethtown College
## 766                 Franklin and Marshall College
## 767                            Gannon University
## 768                               Geneva College
## 769                            Gettysburg College
## 770                            Grove City College
## 771                          Gwynedd Mercy College
## 772                            Haverford College
## 773                          Holy Family College
## 774                            Immaculata College
## 775          Indiana University of Pennsylvania
## 776                               Juniata College
## 777                                 Kings College
## 778                             Lafayette College
## 779                           La Salle University
## 780                        Lebanon Valley College
## 781                             Lehigh University
## 782                            Lincoln University
## 783                              Lycoming College
## 784                              Marywood College
## 785                           Mercyhurst College
## 786                              Messiah College
## 787                             Moravian College
## 788                           Muhlenberg College
## 789                            Widener University
## 790              Bloomsburg Univ. of Pennsylvania
## 791          California University of Pennsylvania
## 792               Edinboro University of Penn.
## 793              Kutztown University of Penn.
## 794      Lock Haven University of Pennsylvania
## 795            Millersville University of Penn.
## 796           Shippensburg University of Penn.
## 797           Slippery Rock University of Penn.
## 798            West Chester University of Penn.
## 799                       University of the Arts
## 800        Philadelphia Coll. of Textiles and Sci.
## 801                           Point Park College
## 802                         Robert Morris College
## 803                            Seton Hill College
## 804                     Saint Josephs University
## 805                        Saint Vincent College
## 806                      Susquehanna University
## 807                           Swarthmore College
## 808                            Temple University
## 809                                 Thiel College
## 810                   University of Pennsylvania
## 811          University of Pittsburgh-Main Campus
## 812                        University of Scranton
## 813                               Ursinus College
## 814                           Villanova University
## 815          Washington and Jefferson College
## 816                             Wilkes University
## 817                                Wilson College
## 818                 York College of Pennsylvania
```

```
## 819               Allentown Coll. of St. Francis de Sales
## 820                               La Roche College
## 821              Pennsylvania State Univ. Main Campus
## 822                Clarion University of Pennsylvania
## 823                               Brown University
## 824                                  Bryant College
## 825                             Providence College
## 826                            Rhode Island College
## 827                   Rhode Island School of Design
## 828                          Salve Regina University
## 829                       University of Rhode Island
## 830                       Roger Williams University
## 831                   Charleston Southern University
## 832                        Central Wesleyan College
## 833                                      The Citadel
## 834                             Clemson University
## 835                                    Coker College
## 836                           College of Charleston
## 837                                  Erskine College
## 838                                Furman University
## 839                                Lander University
## 840                               Newberry College
## 841                            Presbyterian College
## 842            University of South Carolina at Columbia
## 843             University of South Carolina at Aiken
## 844                      Coastal Carolina University
## 845                                Voorhees College
## 846                             Winthrop University
## 847                                  Wofford College
## 848           Univ. of South Carolina at Spartanburg
## 849                         Francis Marion University
## 850                      Dakota Wesleyan University
## 851                          Dakota State University
## 852                             Mount Marty College
## 853                         Northern State University
## 854                              Sioux Falls College
## 855                      South Dakota State University
## 856                       Austin Peay State University
## 857                              Belmont University
## 858                          Carson-Newman College
## 859                      Christian Brothers University
## 860                         David Lipscomb University
## 861               East Tennessee State University
## 862                                   Fisk University
## 863                         Freed-Hardeman University
## 864                                    King College
## 865                              Lambuth University
## 866                                       Lee College
## 867                      Lincoln Memorial University
## 868                               Maryville College
## 869                           University of Memphis
## 870               Middle Tennessee State University
## 871                               Milligan College
## 872                                   Rhodes College
```

```
## 873                      Tennessee Wesleyan College
## 874                             Tusculum College
## 875         University of Tennessee at Chattanooga
## 876           University of Tennessee at Knoxville
## 877              University of Tennessee at Martin
## 878                      University of the South
## 879                        Vanderbilt University
## 880                               Bryan College
## 881                 Abilene Christian University
## 882                    Angelo State University
## 883                             Austin College
## 884                  Concordia Lutheran College
## 885                   Dallas Baptist University
## 886               East Texas Baptist University
## 887                  East Texas State University
## 888                   Hardin-Simmons University
## 889                    Howard Payne University
## 890                 Houston Baptist University
## 891                    Incarnate Word College
## 892                             Lamar University
## 893                      LeTourneau University
## 894                 Lubbock Christian University
## 895                         McMurry University
## 896                 University of North Texas
## 897          Our Lady of the Lake University
## 898                             Rice University
## 899               Sam Houston State University
## 900                           Schreiner College
## 901            Southwest Texas State University
## 902             Southwestern Adventist College
## 903                   Southwestern University
## 904                     St. Edwards University
## 905          St. Marys University of San Antonio
## 906          Stephen F. Austin State University
## 907           Prairie View A. and M. University
## 908                  Tarleton State University
## 909                 Texas Christian University
## 910                  Jarvis Christian College
## 911                               Texas College
## 912                      Texas Lutheran College
## 913                   Texas Southern University
## 914                       Texas Tech University
## 915                  Texas Wesleyan University
## 916                    Texas Womans University
## 917                           Trinity University
## 918                        University of Dallas
## 919            University of Texas at Arlington
## 920               University of Texas at Austin
## 921                 Wayland Baptist University
## 922                 West Texas A&M University
## 923                            Baylor University
## 924          University of Texas at San Antonio
## 925          Texas A&M Univ. at College Station
## 926                         University of Utah
```

```
## 927                          Weber State University
## 928                               Averett College
## 929                             Bluefield College
## 930                    College of William and Mary
## 931                 Christopher Newport University
## 932                     Eastern Mennonite College
## 933                        Emory & Henry College
## 934                                Ferrum College
## 935                     Hampden - Sydney College
## 936                            Hampton University
## 937                               Hollins College
## 938                              Longwood College
## 939                             Lynchburg College
## 940                      James Madison University
## 941                           Mary Baldwin College
## 942                          Marymount University
## 943                       Old Dominion University
## 944                             Radford University
## 945                         Randolph-Macon College
## 946                   Randolph-Macon Womans College
## 947                Virginia Commonwealth University
## 948                               Roanoke College
## 949                          Shenandoah University
## 950                            Sweet Briar College
## 951                         University of Richmond
## 952                       Mary Washington College
## 953   Clinch Valley Coll. of  the Univ. of Virginia
## 954                         George Mason University
## 955                    Virginia Intermont College
## 956                    Virginia Military Institute
## 957                                 Virginia Tech
## 958                    Virginia Union University
## 959                    Virginia Wesleyan College
## 960                  Washington and Lee University
## 961                          University of Virginia
## 962                          Christendom College
## 963                           Bennington College
## 964                       Castleton State College
## 965                        College of St. Joseph
## 966                              Goddard College
## 967                       Green Mountain College
## 968                        Johnson State College
## 969                            Marlboro College
## 970                           Middlebury College
## 971                           Norwich University
## 972                       Saint Michaels College
## 973                        University of Vermont
## 974                 Central Washington University
## 975                 Eastern Washington University
## 976                          Gonzaga University
## 977                 Pacific Lutheran University
## 978                  Seattle Pacific University
## 979                          Seattle University
## 980                         St. Martins College
```

```
## 981                    University of Puget Sound
## 982                    University of Washington
## 983                 Washington State University
## 984                Western Washington University
## 985                             Whitman College
## 986                           Whitworth College
## 987                     Evergreen State College
## 988                             Alverno College
## 989                              Beloit College
## 990                            Carthage College
## 991               Concordia University Wisconsin
## 992                            Lakeland College
## 993                         Lawrence University
## 994                 Marian College of Fond du Lac
## 995                         Marquette University
## 996                Milwaukee School of Engineering
## 997                          Mount Mary College
## 998                           Northland College
## 999                               Ripon College
## 1000                         St. Norbert College
## 1001          University of Wisconsin at Madison
## 1002         University of Wisconsin at Milwaukee
## 1003         University of Wisconsin at Green Bay
## 1004                             Viterbo College
## 1005              University of Wisconsin-Stout
## 1006          Univ. of Wisconsin at Eau Claire
## 1007          Univ. of Wisconsin at Platteville
## 1008            University of Wisconsin-Superior
## 1009          University of Wisconsin-Whitewater
## 1010            Univ. of Wisconsin at OshKosh
## 1011              Wisconsin Lutheran College
## 1012               Alderson-Broaddus College
## 1013                 Bluefield State College
## 1014                           Concord College
## 1015                    Davis & Elkins College
## 1016                    Fairmont State College
## 1017                   Glenville State College
## 1018                        Marshall University
## 1019                 University of Charleston
## 1020                    Salem-Teikyo University
## 1021                          Shepherd College
## 1022               West Liberty State College
## 1023                 West Virginia University
## 1024          West Virginia Wesleyan College
## 1025                   Wheeling Jesuit College
## 1026                      University of Wyoming
```

```r
# college_data %>% select(State) %>% distinct() # get distinct values for
# state column OR
nrow(college_data)
```

```
## [1] 1042
```

```r
length(levels(college_data$Name))
```

`## [1] 1026`

1026 distinct colleges represented.

There are 1042 colleges. Note that there are 1026 different names of colleges - these are colleges with same name in different states. There are two categorical variables in this dataset: `Name` and `State`.

How many variables?

```r
13
```

`## [1] 13`

List all variable names

```
##  [1] "Name"           "State"           "Schooltype"
##  [4] "All.test.std"   "App.accept"      "Acc.Rate"
##  [7] "Pct.Yield"      "Total.students"  "Student.Faculty"
## [10] "Grad.rate"      "Pct.fac.degree"  "In.Tuition"
## [13] "Room.board"
```

continuous vs. categorical variables

???? `Name` and `State`

b) Which school has the highest graduation rate, and what is that rate? Which school has the lowest graduation rate, and what is that rate? What was Penn's graduation rate in 1995? What is the mean graduation rate across all schools?

Which school has the highest graduation rate

```r
# max(college_data$Grad.rate)
# college_data$name[which.max(college_data$Grad.rate)]
college_data %>% select(Name, Grad.rate) %>% filter(Grad.rate == max(Grad.rate))
```

```
##                         Name Grad.rate
## 1          Harvey Mudd College       100
## 2        Santa Clara University       100
## 3              Amherst College       100
## 4            Harvard University       100
## 5            Lindenwood College       100
## 6                Siena College       100
## 7    College of Mount St. Joseph       100
## 8            Grove City College       100
## 9        University of Richmond       100
## 10             Goddard College       100
```

```r
# college_data[which(college_data$Grad.rate ==
# max(college_data$Grad.rate)),] %>% select(Name, Grad.rate)
```

There are 10 schools with 100% graduation rate. (It's ok if you only pick up one.)

Which school has the lowest graduation rate, and what is that rate?

```
college_data %>% select(Name, Grad.rate) %>% filter(Grad.rate == min(Grad.rate))
```

```
##                          Name Grad.rate
## 1 Texas Southern University        10
```

The lowest graduation rate belongs to Texas Southern University at 10%.

What was Penn's graduation rate in 1995?

```
Penn
```

```
##                           Name State Schooltype All.test.std App.accept
## 820 University of Pennsylvania    PA          2     2.423741       5232
##      Acc.Rate Pct.Yield Total.students Student.Faculty Grad.rate
## 820 42.21397   47.0948           9736             6.3        93
##      Pct.fac.degree In.Tuition Room.board
## 820             96      17020       7270
```

```
college_data %>% select(Name, Grad.rate) %>% filter(Name == "University of Pennsylvania")
```

```
##                         Name Grad.rate
## 1 University of Pennsylvania        93
```

```
# OR college_data %>% filter(Name == 'University of Pennsylvania') %>%
# select(Name, Grad.rate)
```

Penn's graduation rate:

```
college_data %>% filter(Name == "University of Pennsylvania") %>% select(Name,
    Grad.rate)
```

```
##                         Name Grad.rate
## 1 University of Pennsylvania        93
```


What is the mean graduation rate across all schools?

```
mean(college_data$Grad.rate)
```

```
## [1] 61.77543
```

```
# OR college_data %>% summarize(mean_grad = mean(Grad.rate))
```

61.77543

c) Give the histogram of graduation rate and write a very short summary (max 3 sentences) about this distribution.

```
ggplot(college_data) + geom_histogram(aes(x = Grad.rate), bins = 50, fill = "blue") +
    labs(title = "Histogram of the graduation rates across universities", x = "Graduation rate",
        y = "Frequency")
```

Histogram of the graduation rates across universities
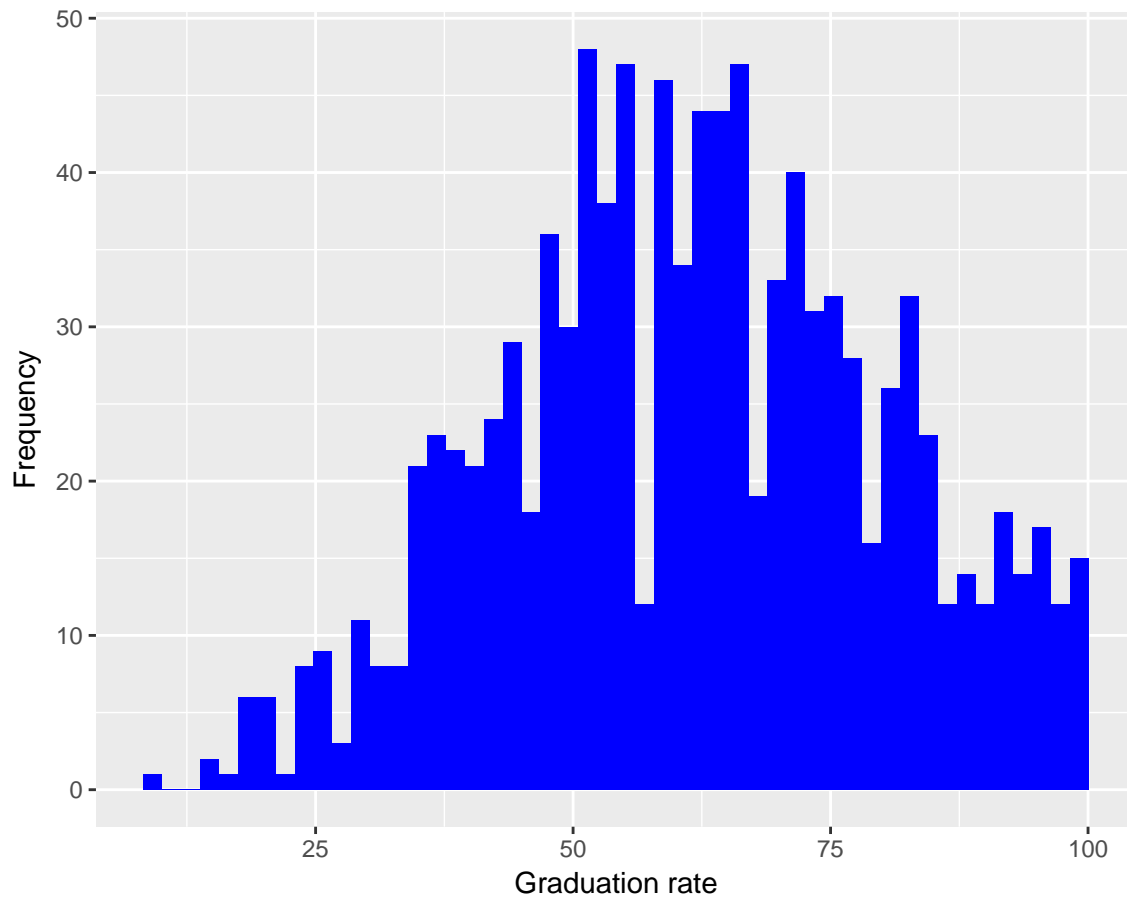
```
# OR hist(college_data$Grad.rate, breaks = 10)

# OR college_data %>% ggplot(aes(Grad.rate)) + geom_histogram(bins = 90) +
# xlab('Graduation Rate') + ylab('Count') + ggtitle('Graduation Rate
# Histogram') + theme_bw()
```

Assume all linear model assumptions are met in the following analyses.

Question 2: School Type and State vs Graduation Rate

a) Make side to side boxplots of **graduation rate** vs **school type**. Does one type seems to have higher a graduation rate compared to the other? Write a short summary (max 3 sentences) of this finding. Does that agree with your intuition about private schools (`Schooltype = 2`) vs. public schools (`Schooltype = 1`)?

```
# names(college_data)
college_data %>% ggplot(aes(x = Schooltype, y = Grad.rate)) + geom_boxplot() +
    theme(axis.text.x = element_text(angle = 30, hjust = 1)) + labs(title = "Boxplots of school type ver
    x = "School type (1-public, 2-private)", y = "graduation rate")
```

Boxplots of school type versus graduation rate
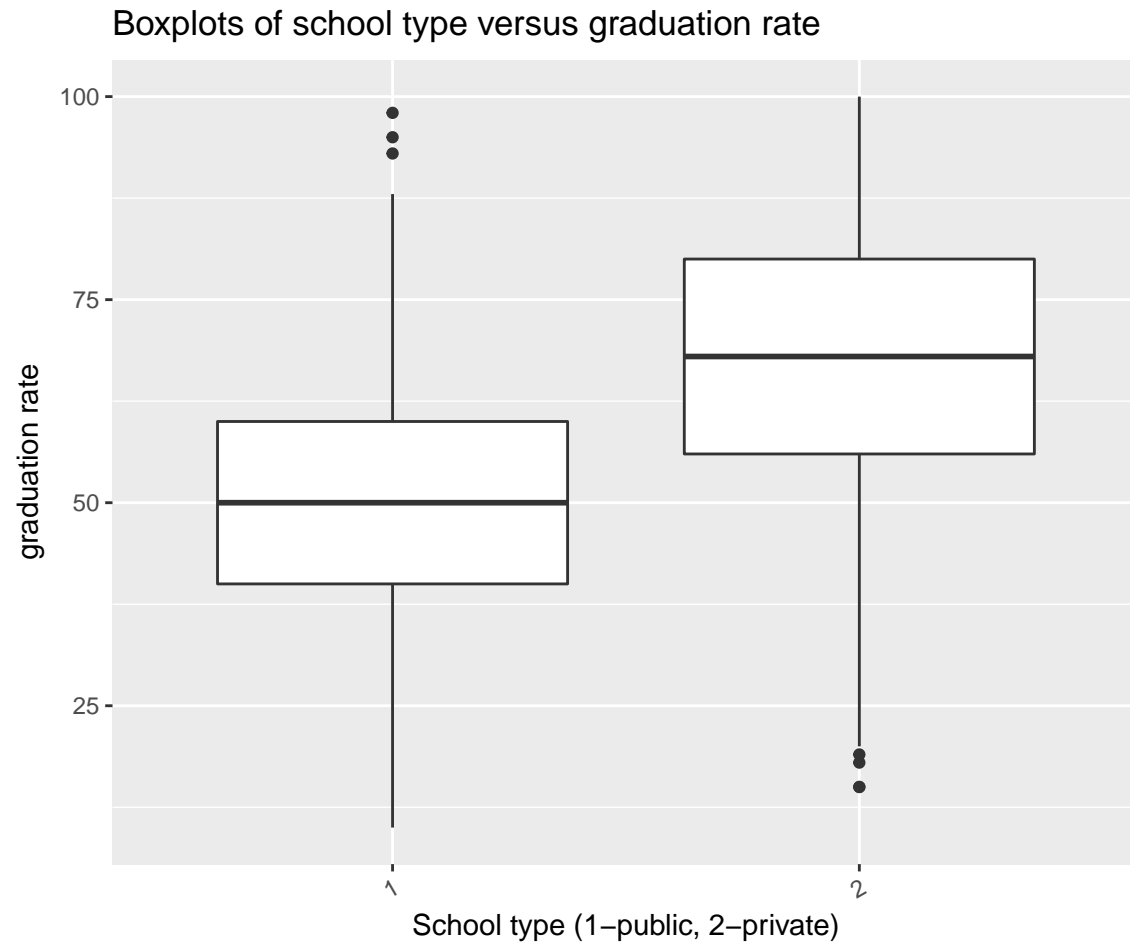

```
# OR boxplot(college_data$Grad.rate ~ college_data$Schooltype)
```

Public schools on left, private schools on right. Private schools appear to have higher graduation rates.

b) `fit1`: `Grad.rate` vs. `Schooltype`

Perform a test to determine if the mean `Grad.rate` between the two school types is different at 0.01 level. Which type has a higher `Grad.rate`? Produce a 95% confidence interval for the mean difference.

We can perform a t-test:

```
private_grad_rates <- college_data %>% filter(Schooltype == 2) %>% select(Grad.rate) %>%
    unlist
public_grad_rates <- college_data %>% filter(Schooltype == 1) %>% select(Grad.rate) %>%
    unlist
t.test(public_grad_rates, private_grad_rates)
```

```
##
##  Welch Two Sample t-test
##
## data:  public_grad_rates and private_grad_rates
## t = -15.969, df = 788.87, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.88787 -14.75274
## sample estimates:
```

```
## mean of x mean of y
##  50.62108  67.44139
```

Clarify t-test assumption of normality - looks fine.

```
par(mfrow = c(1, 2))
hist(private_grad_rates)
hist(public_grad_rates)
```

Histogram of private_grad_rate **Histogram of public_grad_rate**
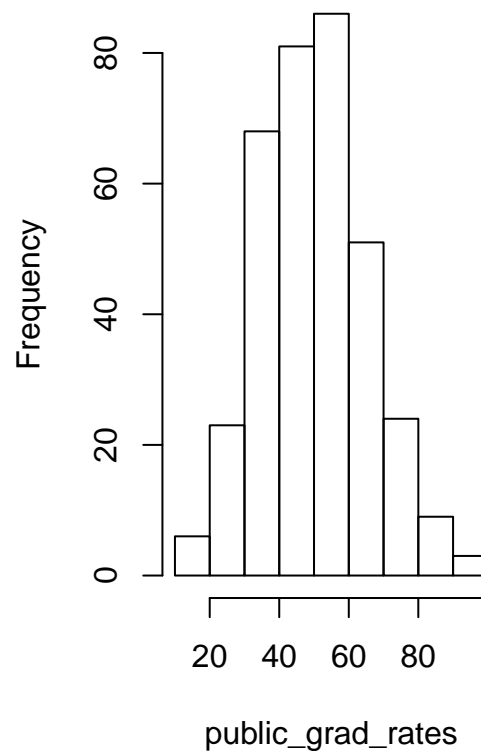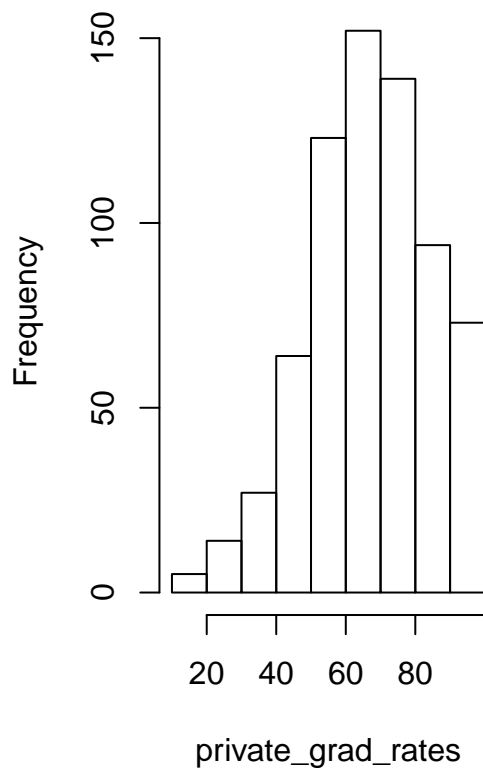
Then, we can back out the confidence interval via the above summary output:

```
95 percent confidence interval:
 -18.88787 -14.75274
```

You can also do a regression based approach, this is probably the easier one. Note that the results are the same!

```
fit1 <- lm(Grad.rate ~ Schooltype, data = college_data)
confint(fit1, parm = 2)
```

```
##                 2.5 %   97.5 %
## Schooltype2 14.66475 18.97587
```

c) `fit1.1: Grad.rate` vs. `State`

Can we prove that the mean graduation rates are different, at the 0.01 level, among all the states? Which state appears to have the highest graduation rate, and which state appears to have the lowest graduation

rate? Note that according to standard R-coding `AK`/Alaska is the base case in this analysis.

Some states have high p-values, that is, they are not significantly different from Alaska, the base case.

```
fit1.1 <- lm(Grad.rate ~ State, data = college_data)
summary(fit1.1) %>% broom::tidy() %>% arrange(-p.value) %>% head
```

```
##            term estimate std.error statistic   p.value
## 1 (Intercept)  15.0000  16.76279 0.8948389 0.3710905
## 2      StateUT  22.5000  20.53015 1.0959493 0.2733672
## 3      StateWY  30.0000  23.70617 1.2654932 0.2059920
## 4      StateNV  26.5000  20.53015 1.2907847 0.1970794
## 5      StateHI  27.5000  20.53015 1.3394936 0.1807171
## 6      StateLA  25.9375  17.27869 1.5011262 0.1336414
```

Lowest grad rate states:

```
college_data %>% group_by(State) %>% summarize(grad_rate = mean(Grad.rate)) %>%
    arrange(grad_rate) %>% head
```

```
## # A tibble: 6 x 2
##     State grad_rate
##    <fctr>     <dbl>
## 1      AK  15.00000
## 2      UT  37.50000
## 3      LA  40.93750
## 4      NV  41.50000
## 5      HI  42.50000
## 6      OK  42.61538
```

Highest grad rate states:

```
college_data %>% group_by(State) %>% summarize(grad_rate = mean(Grad.rate)) %>%
    arrange(-grad_rate) %>% head
```

```
## # A tibble: 6 x 2
##     State grad_rate
##    <fctr>     <dbl>
## 1      RI  76.62500
## 2      DC  75.50000
## 3      PA  74.35065
## 4      CT  74.06250
## 5      MA  73.87500
## 6      VT  73.66667
```

d) `fit1.2`: `Grad.rate` vs. `Schooltype` and `State`

Controlling the school type, is the state where the school locates a useful factor at the .01 level?

Yes, as shown clearly in the Anova output, State is significant at under the 0.01 level.

```
fit1.2 <- lm(Grad.rate ~ Schooltype + State, data = college_data)
car::Anova(fit1.2)
```

```
## Anova Table (Type II tests)
##
## Response: Grad.rate
##             Sum Sq  Df F value    Pr(>F)
## Schooltype   41957   1 175.628 < 2.2e-16 ***
## State        55615  50   4.656 < 2.2e-16 ***
```

```
## Residuals  236506 990
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 3: Faculty Effects

The variable `Pct.fac.degree` summarizes the percentage of faculty members who hold higher education degrees.

Construct `fit2`: `Grad.rate` vs. `Pct.fac.degree`

a) Report the summary of your linear model. Is `Pct.fac.degree` a significant variable in this model at .05 level? How does `Pct.fac.degree` affect `Grad.rate`?

`Pct.fac.degree` is a significant variable in this model at .05 level. For every percent of faculty members who have a higher degree, it increases graduation rate by 0.34 percent.

```
fit2 <- lm(Grad.rate ~ Pct.fac.degree, data = college_data)
summary(fit2)
```

```
##
## Call:
## lm(formula = Grad.rate ~ Pct.fac.degree, data = college_data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -50.87 -12.00   0.21  12.47  54.68
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    35.05981    2.69889   12.99   <2e-16 ***
## Pct.fac.degree  0.34418    0.03405   10.11   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.7 on 1040 degrees of freedom
## Multiple R-squared:  0.08948,	Adjusted R-squared:  0.08861
## F-statistic: 102.2 on 1 and 1040 DF,  p-value: < 2.2e-16
```

b) Make a scatter plot with $y = $ `Grad.rate` and $x = $ `Pct.fac.degree`. Overlay `fit2` onto the plot.

```
college_data %>% ggplot(aes(x = Pct.fac.degree, y = Grad.rate)) + geom_point() +
    geom_abline(slope = 0.334418, intercept = 35.05981) + xlab("% Faculty w/Higher Degree") +
    ylab("Graduation Rate") + theme_bw()
```
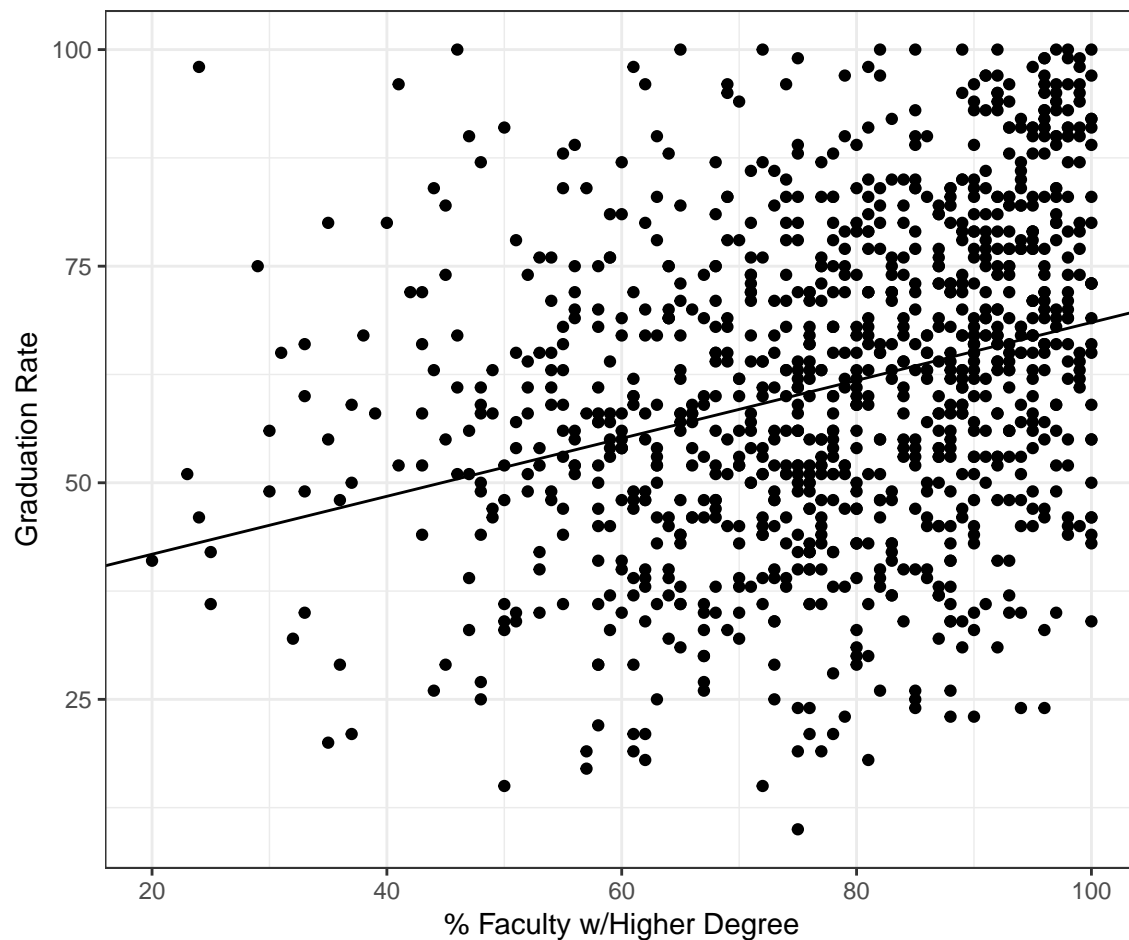
Construct `fit2.1`: `Grad.rate` vs. `Pct.fac.degree + All.test.std`

```
fit2.1 <- lm(Grad.rate ~ Pct.fac.degree + All.test.std, data = college_data)
```

c) Is `Pct.fac.degree` still a significant variable in this model at the .05 level?

Yes

```
summary(fit2.1)
```

```
##
## Call:
## lm(formula = Grad.rate ~ Pct.fac.degree + All.test.std, data = college_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -44.016  -9.744  -0.452   9.181  46.204
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    68.74170    2.69286  25.527   <2e-16 ***
## Pct.fac.degree -0.08538    0.03409  -2.505   0.0124 *
## All.test.std   12.29484    0.55413  22.188   <2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 14.59 on 1039 degrees of freedom
## Multiple R-squared:  0.3822, Adjusted R-squared:  0.381
## F-statistic: 321.4 on 2 and 1039 DF,  p-value: < 2.2e-16
```

d) Interpret the coefficent of `Pct.fac.degree` in `fit2.1`.

This model suggests that it's worse for higher % of faculty members to have higher education degrees.

e) Why might the two estimates of beta for `Pct.fac.degree` differ?

This new model controls for the 'quality' of students admitted. The two variables under consideration have a high correlation of 0.5679462, a condition known as collinearity.

Question 4: Parsimonious Models (all subsets)

Construct `fit3`: a model with all available sensible variables

```
fit3 <- lm(Grad.rate ~ . - Name, data = college_data)
```

Based on fit3, answer the following questions:

a) Is `State` a significant variable at .01 level after controlling for all other variables in the model? Provide an appropriate test.

We can do Anova test - State is significant.

```
car::Anova(fit3)
```

```
## Anova Table (Type II tests)
##
## Response: Grad.rate
##                 Sum Sq  Df  F value    Pr(>F)
## State            28575  50   3.9148 < 2.2e-16 ***
## Schooltype        3891   1  26.6535 2.946e-07 ***
## All.test.std     27381   1 187.5645 < 2.2e-16 ***
## App.accept        3919   1  26.8483 2.671e-07 ***
## Acc.Rate           473   1   3.2414    0.0721 .
## Pct.Yield           40   1   0.2754    0.5999
## Total.students    2613   1  17.8994 2.547e-05 ***
## Student.Faculty     24   1   0.1674    0.6825
## Pct.fac.degree      27   1   0.1848    0.6674
## In.Tuition           0   1   0.0004    0.9840
## Room.board         253   1   1.7314    0.1885
## Residuals       143209 981
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b) If you want to kick one variable out from this model such that the resulting model would have the smallest possible RSS, which variable would you choose, and why?

Kick out instate tuition - variable with smallest F-value / least importance.

Remove `State` from the data under consideration but include all other variables. Construct `fit4`: a parsimonious model, using regusubsets with exhaustive search.

```
colleges_sub <- college_data %>% select(-Name, -State)
x_col <- model.matrix(Grad.rate ~ ., colleges_sub)[, -1]
y_col <- college_data$Grad.rate
reg_search <- regsubsets(x_col, y_col, method = "exhaustive")
```

c) Show the Cp plot and also show the BIC plot. Based on the two plots, which is the most desirable model size? Why?

```
par(mfrow = c(1, 2))
plot(summary(reg_search)$bic, main = "BIC Plot")
plot(summary(reg_search)$cp, main = "CP Plot")
```
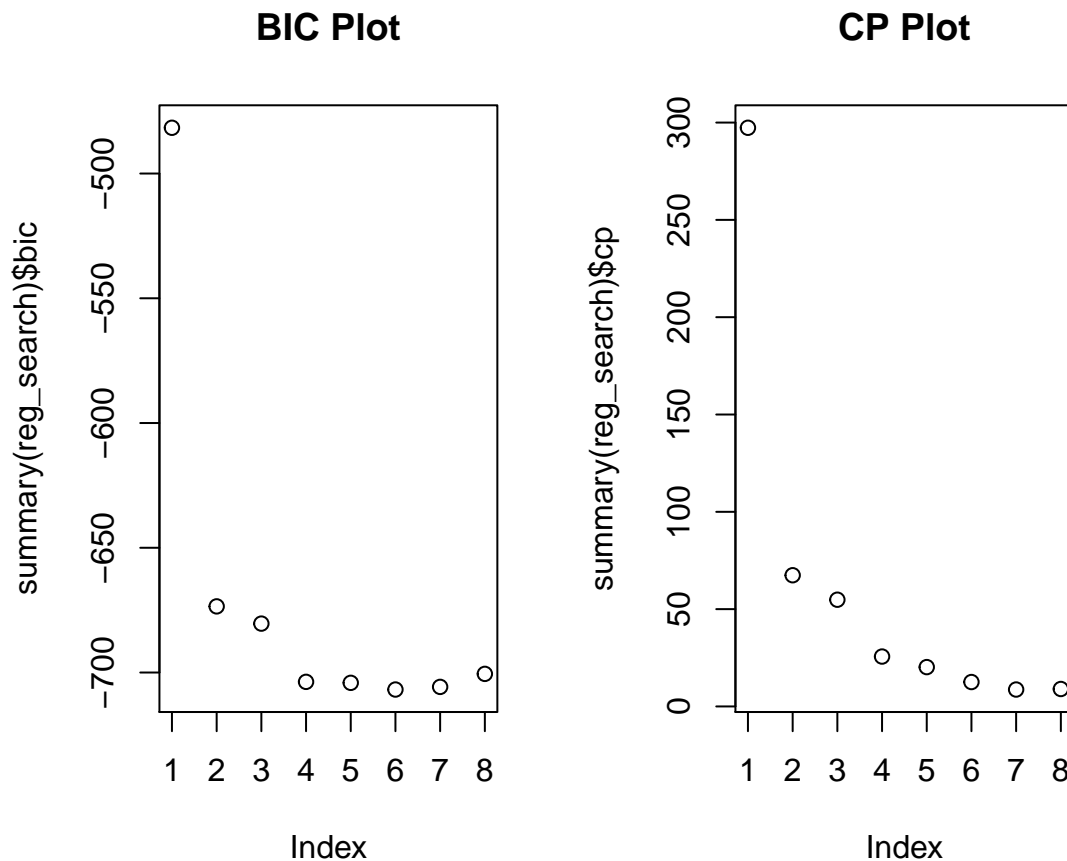


```
# find min via BIC: which(summary(reg_search)$bic ==
# min(summary(reg_search)$bic)) find min via CP:
# which(summary(reg_search)$cp == min(summary(reg_search)$cp))
```

d) Regardless of your answer in c), report the 4-variable model chosen by regsubsets. To save time we will not pursue further.

```
All.test.std, App.accept, Total.students, In.tuition
```

Question 5: Parsimonious Models (LASSO)

Use LASSO for model selection, again making sure to do so without including the `State` variable in the LASSO process.

a) Run `cv.glmnet()` with `set.seed(12)`. Plot `cmv` vs. `lambda`.

```
set.seed(12)
fit_lasso <- cv.glmnet(x_col, y_col, family = "gaussian", alpha = 1)
plot(fit_lasso, main = "CMV vs Lambda in LASSO")
```
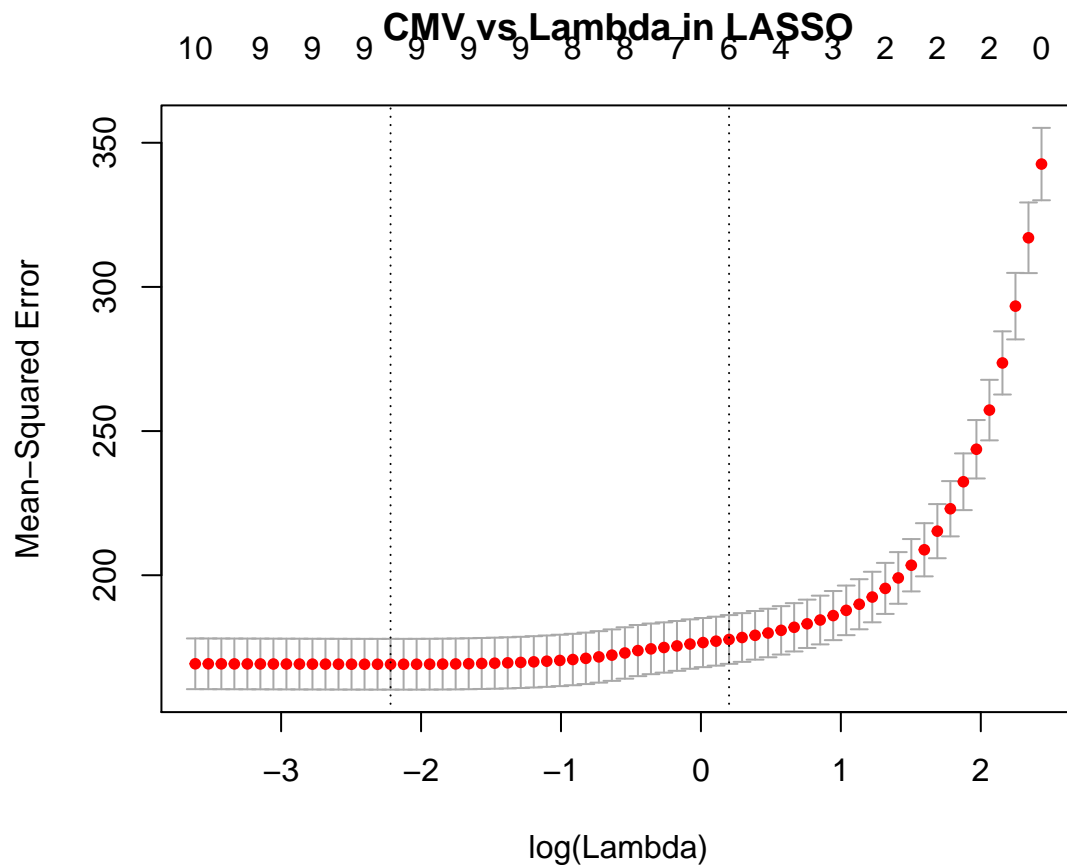

b) What is the `lambda.1se` value? Under the `lambda.1se` criterion, list the non-zero variables returned?

lambda 1se:

```
fit_lasso$lambda.1se
```

```
## [1] 1.222679
```

And we have the following variables in the model:

```
coefs_1se = coef(fit_lasso, s = "lambda.1se")
rownames(coefs_1se)[which((coefs_1se) != 0)]
```

```
## [1] "(Intercept)"  "Schooltype2"  "All.test.std" "Acc.Rate"
## [5] "Pct.Yield"    "In.Tuition"   "Room.board"
```

c) `fit5`: Run OLS with all the variables returned from part b), **and with State also included in the model**. Are all the variables included here significant at the .01 level? If not, perform backward elimination (manually) until all the p-values for the remaining variables are less than .01. Show your model building process and report the final LS equations. *Note: for this problem, force State into the final model, i.e., do not remove State.*

```
fit5 = lm(Grad.rate ~ State + Schooltype + All.test.std + Acc.Rate + Pct.Yield +
    In.Tuition + Room.board, data = college_data)

Anova(fit5)

# remove In.Tuition
fit5_1 = update(fit5, . ~ . - In.Tuition)
Anova(fit5_1)

# remove Room.Board
fit5_2 = update(fit5_1, . ~ . - Room.board)
Anova(fit5_2)

# remove Pct.Yield
fit5_3 = update(fit5_2, . ~ . - Pct.Yield)
Anova(fit5_3)

# remove Acc.Rate
fit5_4 = update(fit5_3, . ~ . - Acc.Rate)
Anova(fit5_4)
```

We see that our final model is `Grad.rate ~ State + Schooltype + All.test.std` :)

Question 6: Graduation Evaluation

Independent of Question 5. Assume that this we've decided to use `fit6` as our final model.

fit6: Grad.rate ~ State + Schooltype + All.test.std

`fit6 <- lm(Grad.rate ~ State + Schooltype + All.test.std, data = college_data)`

a) Are all three variables significant at .01 level?

Yep

```
car::Anova(fit6)
```

```
## Anova Table (Type II tests)
##
## Response: Grad.rate
##                Sum Sq  Df  F value    Pr(>F)
## State           42005  50   5.6109 < 2.2e-16 ***
## Schooltype      20239   1 135.1712 < 2.2e-16 ***
## All.test.std    88427   1 590.5889 < 2.2e-16 ***
## Residuals      148079 989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
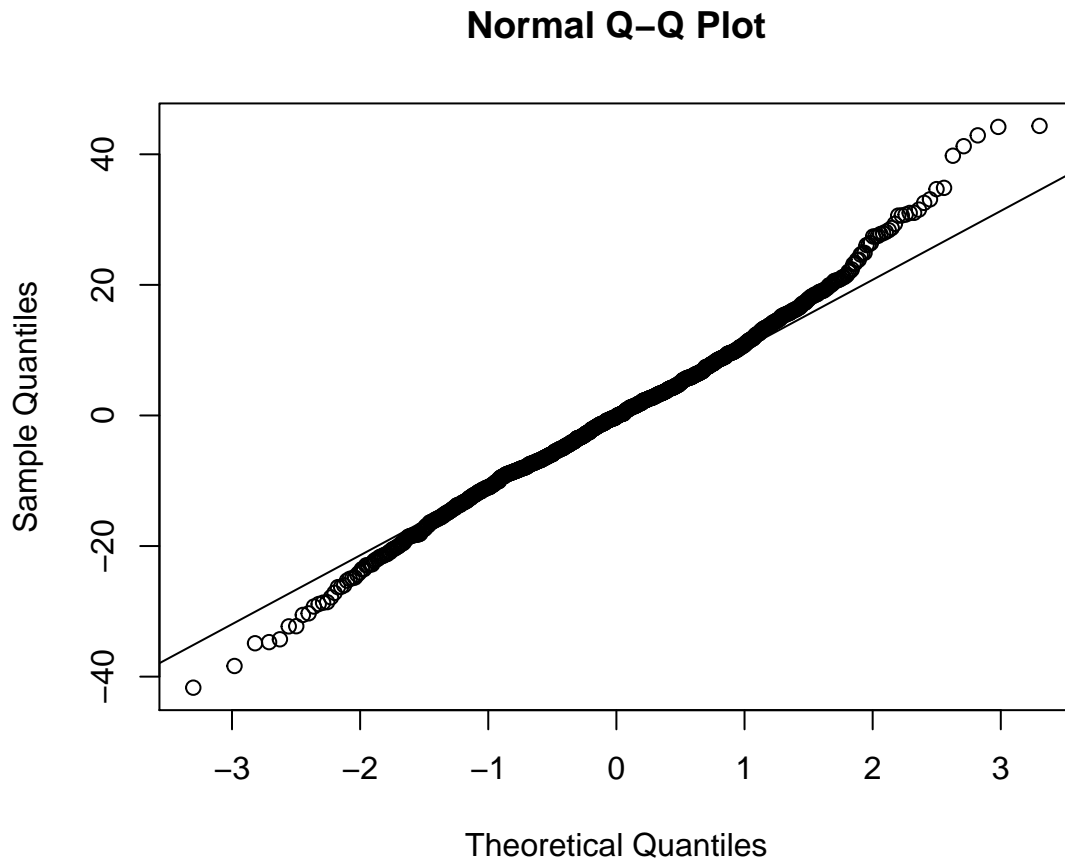
b) Provide the residual plot.

```
plot(fit6$fitted, fit6$residuals)
```
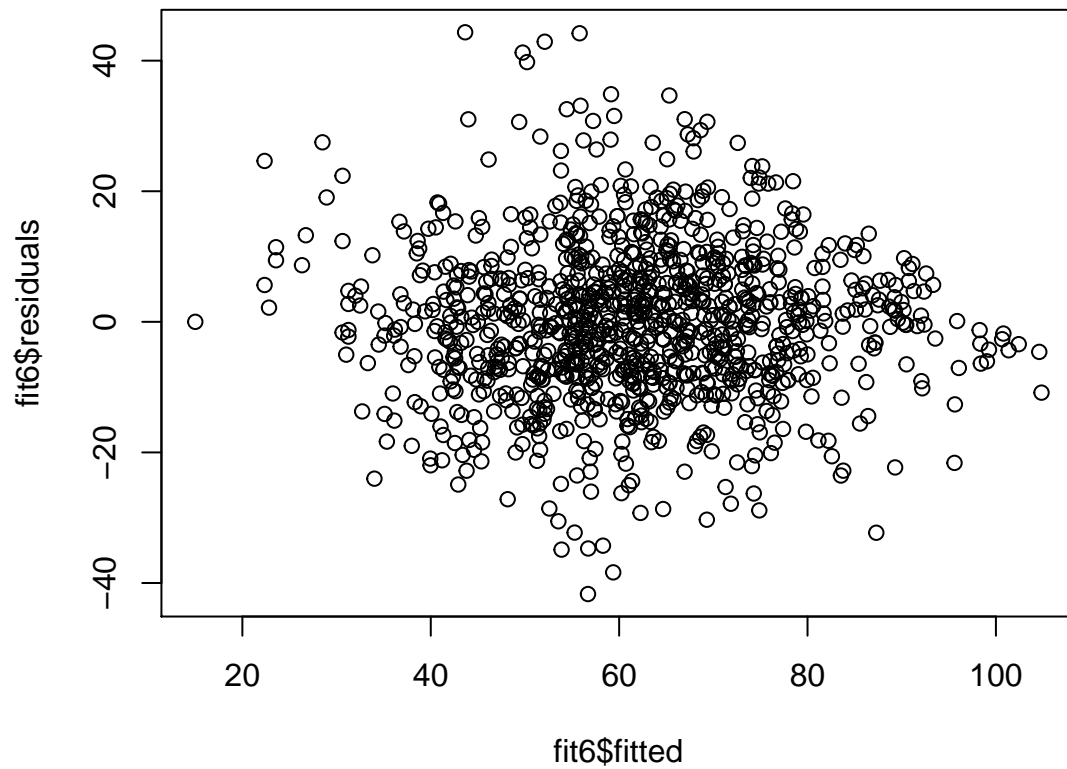

c) Provide the qqnorm plot of the residuals.

```
qqnorm(fit6$residuals)
qqline(fit6$residuals)
```

Normal Q–Q Plot

d) Do the model meet all the linear model assumptions?

Yeah

e) Finally, using `fit6`, provide a 95% prediction interval for Penn's graduation rate. Based on Penn's actual graduation rate, how do you think of the performance of our prediction?

```
predict(fit6, newdata = filter(college_data, Name == "University of Pennsylvania"),
    interval = "prediction", level = 0.95)
```

```
##        fit      lwr      upr
## 1 99.04858 74.81411 123.2831
```

The prediction interval is 74.81 to 123%, which I guess isn't great because our upper bound is over 100%, but is acceptable. Our fit is at 99%, and Penn's real graduation rate is 93, indicating that this school is underperforming our model, but at least taught us enough statistics to get a prediction interval containing the true rate.

Question 7: Freedom of the Press

Newsweek did a great job of collecting graduate data, but some schools are unhappy with their exact graduation figures being reported. They lobbied Newsweek's publisher to report only whether a school's graduation rate is either High (`Grad.rate >= 70`) or Low (`Grad.rate < 70`); the "journalists" acquiesced

to their corporate overlords. From now on, the only graduation rate data available to you is in that high/low form.

a) Create a new categorical variable `Grad.rate.2` in `college_data` that fits the new specification. What proportion of the schools are categorized as "High Graduation", that is, `Grad.rate.2 == "1"`?

34.06% of colleges have "high" graduation rates.

```
college_data %<>% mutate(Grad.rate.2 = as.factor(as.numeric(Grad.rate >= 70)))
table(college_data$Grad.rate.2)/nrow(college_data)
```

```
##
##        0        1
## 0.659309 0.340691
```

b) How well can we predict `Grad.rate.2`, with only three variables: `State`, `Schooltype` and `All.test.std`. Run a logistic regression of `Grad.rate.2` vs. `State`, `Schooltype` and `All.test.std`. Is every variable significant at .01 level, whilst controlling the other two variables?

Each variable is significant at 0.01, while controlling for the other variables.

```
fit7 <- glm(Grad.rate.2 ~ State + Schooltype + All.test.std, data = college_data,
    family = "binomial")
car::Anova(fit7)
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: Grad.rate.2
##              LR Chisq Df Pr(>Chisq)
## State         141.269 50  1.212e-10 ***
## Schooltype     71.555  1  < 2.2e-16 ***
## All.test.std  200.264  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c) Let us fix our classification threshold to 0.5, that is, we will classify the school to be "High Graduation" if $\hat{P}($`Grad.rate.2 == "1"`$) > 0.5$ (the estimated probability of being "High Graduation" is greater than 0.5). Under this framework, what is the in-sample mis-classification error? Show your work.

```
con_mat <- table(fit7$fitted.values > 0.5, college_data$Grad.rate.2)

1 - sum(diag(con_mat))/nrow(college_data)
```

```
## [1] 0.2072937
```

```
# sum of matrix diagonals is # of correctly predicted values 1 - correct =
# MCE
```

END