

DATA 607: Project 2

James Williams

Import Libraries

```
library(tidyr)
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
```

```
library(stringr)
library(ggplot2)
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##   src, summarize

## The following objects are masked from 'package:base':
##
##   format.pval, units
```

DATA SET 1: National Inventory of Dams: Arkansas

<https://nid.sec.usace.army.mil/ords/f?p=105:19:10737731657977::NO::>

```
dams <- read.csv("https://raw.githubusercontent.com/jambawilliams/DATA607_PROJECT2/master/AK_U.csv", header = TRUE)
head(dams)
```

```
## RECORDID DAM_NAME OTHER_DAM_NAME DAM_FORMER_NAME STATEID NIDID
## 1 1 COOPER LAKE NA AK00001
## 2 2 BLUE LAKE NA AK00002
## 3 3 SALMON CREEK NA AK00003
## 4 4 ANNEX CREEK NA AK00004
## 5 5 CRYSTAL LAKE NA AK00005
## 6 6 KETCHIKAN LAKES NA AK00006
## LONGITUDE LATITUDE SECTION COUNTY
## 1 -149.8231 60.43371 N/A KENAI PENINSULA
## 2 -135.1917 57.06330 T55S, R64E, S35 SITKA
## 3 -134.4036 58.34185 T41S,R67E,S2 JUNEAU
## 4 -134.1266 58.32694 T41S,R69E,S9 JUNEAU
## 5 -132.8455 56.60000 T61S,R81E,S6 WRANGELL-PETERSBURG
## 6 -131.6203 55.35940 T75S,R90E,S17 KETCHIKAN GATEWAY
## RIVER CITY DISTANCE
## 1 COOPER CREEK, KENAI RIVER COOPER LANDING 6
## 2 SAWMILL CREEK NONE 0
## 3 SALMON CREEK, GASTINEAU INLET JUNEAU 2
## 4 ANNEX CREEK, TAKU INLET NONE 1
## 5 CRYSTAL CREEK, BLIND RIVER N/A 0
## 6 KETCHIKAN CREEK KETCHIKAN, AK 2
## OWNER_NAME OWNER_TYPE DAM_DESIGNER PRIVATE_DAM
## 1 CHUGACH ELECTRIC ASSOCIATION, INC. P CAREY AND KRAMER Y
## 2 CITY AND BOROUGH OF SITKA L CAREY AND KRAMER N
## 3 ALASKA ELECTRIC LIGHT & POWER COMPANY P LARS JORGENSEN Y
## 4 ALASKA ELECTRIC LIGHT & POWER COMPANY P FELIX J. TONER Y
## 5 CITY OF PETERSBURG L N
## 6 CITY OF KETCHIKAN L N
## DAM_TYPE CORE FOUNDATION PURPOSES YEAR_COMPLETED YEAR_MODIFIED DAM_LENGTH
## 1 RE FCZ RSZ H 1959 920
## 2 VA HCK RK HRS 1961 288
## 3 VA HCK RK HORS 1914 648
## 4 OT FXK RK HR 1968 115
## 5 ER FMK K H 1955 205
## 6 ER IXK RK HS 1957 1163
## DAM_HEIGHT STRUCTURAL_HEIGHT HYDRAULIC_HEIGHT NID_HEIGHT MAX_DISCHARGE
## 1 52 52.0 39 52 24500
## 2 294 210.0 136 294 14000
## 3 170 175.0 140 175 810
## 4 20 20.0 20 20 458
## 5 43 34.5 30 43 2632
## 6 22 34.0 27 34 7500
## MAX_STORAGE NORMAL_STORAGE NID_STORAGE SURFACE_AREA DRAINAGE_AREA HAZARD EAP
## 1 127600 112000 127600 3100 31.0 H Y
## 2 266000 266000 266000 1646 37.0 H Y
## 3 18000 12000 18000 165 5.0 H Y
```

## 4	23400	23400	23400	264	6.0	L	NR
## 5	5800	5200	5800	233	1.5	H	Y
## 6	13353	11463	13353	55	8.1	H	Y
##	INSPECTION_DATE	INSPECTION_FREQUENCY	STATE_REG_DAM	STATE_REG_AGENCY			
## 1	7/27/2017		1	N			
## 2	5/15/2018		1	N			
## 3	5/17/2018		1	N			
## 4	5/16/2018		3	N			
## 5	8/31/2017		1	N			
## 6	6/19/2018		1	N			
##	SPILLWAY_TYPE	SPILLWAY_WIDTH	OUTLET_GATES	VOLUME	NUMBER_OF_LOCKS		
## 1	U	50		0	0		
## 2	U	140		0	0		
## 3	U	60		52000	0		
## 4	U	57		0	0		
## 5	U	46		0	0		
## 6	U	0		0	0		
##	LENGTH_OF_LOCKS	WIDTH_OF_LOCKS	FED_FUNDING	FED_DESIGN	FED_CONSTRUCTION		
## 1	0	0					
## 2	0	0					
## 3	0	0					
## 4	0	0					
## 5	0	0					
## 6	0	0					
##	FED_REGULATORY	FED_INSPECTION	FED_OPERATION	FED_OWNER	FED_OTHER	SOURCE_AGENCY	
## 1	FERC	FERC				FERC	
## 2	FERC	FERC				FERC	
## 3	FERC	FERC				FERC	
## 4	FERC	FERC				FERC	
## 5	FERC	FERC				FERC	
## 6	FERC	FERC				FERC	
##	STATE	SUBMIT_DATE	URL_ADDRESS	CONG_NAME	PARTY	CONG_DIST	OTHERSTRUCTUREID
## 1	AK	05-AUG-18	NA Don Young (R)	R	AK00		
## 2	AK	05-AUG-18	NA Don Young (R)	R	AK00		
## 3	AK	05-AUG-18	NA Don Young (R)	R	AK00		
## 4	AK	05-AUG-18	NA Don Young (R)	R	AK00		
## 5	AK	05-AUG-18	NA Don Young (R)	R	AK00		
## 6	AK	05-AUG-18	NA Don Young (R)	R	AK00		
##	NUMSEPARATESTRUCTURES	PERMITTINGAUTHORITY	INSPECTIONAUTHORITY				
## 1	1	N	N				
## 2	1	N	N				
## 3	1	N	N				
## 4	2	N	N				
## 5	1	N	N				
## 6	1	N	N				
##	ENFORCEMENTAUTHORITY	JURISDICTIONALDAM	EAP_LAST_REV_DATE				
## 1	N	N	08-JAN-18				
## 2	N	N	27-DEC-17				
## 3	N	N	27-SEP-17				
## 4	N	N	27-SEP-17				
## 5	N	N	19-DEC-17				
## 6	N	N	04-JAN-18				

Clean data

Remove excess columns, null values, and empty cells

```
dams <- dams[, c("DAM_NAME", "CITY", "YEAR_COMPLETED", "MAX_STORAGE")]
dams %>%
  mutate_all(~ifelse(. %in% c("N/A", "NONE", "null", ""), NA, .)) %>%
  na.omit()
```

	DAM_NAME	CITY	YEAR_COMPLETED
## 1	COOPER LAKE	COOPER LANDING	1959
## 3	SALMON CREEK	JUNEAU	1914
## 6	KETCHIKAN LAKES	KETCHIKAN, AK	1957
## 11	CARLANNA LAKE	KETCHIKAN	1975
## 13	WRANGELL UPPER	WRANGELL	1967
## 14	WRANGELL LOWER	WRANGELL	1968
## 15	PETERSBURG UPPER DAM	PETERSBURG	1950
## 16	HESS CREEK DAM	LIVENGOD	1946
## 17	MEALS LAKE DAM	CORDOVA	1972
## 18	PILLAR CREEK DAM NO.1.A	KODIAK	1969
## 19	PILLAR CREEK DAM NO. 2.C	KODIAK	1968
## 20	BETTINGER UPPER RESERVOIR DAM	KODIAK	1968
## 21	SELDOVIA UPPER DAM	SELDOVIA	1953
## 24	CAMPBELL LAKE DAM	ANCHORAGE	1957
## 25	WESTCHESTER LAGOON DAM	ANCHORAGE	1973
## 27	SHOTTER CREEK UPPER DAM	HOONAH	1968
## 28	EKLUTNA DAM	EKLUTNA VILLAGE	1965
## 29	LAKE "O" THE HILLS	ANCHORAGE	1975
## 30	SHIP CREEK DAM	ANCHORAGE	1954
## 34	LONG LAKE DAM	JUNEAU	1973
## 35	EYAK LAKE	CORDOVA	1972
## 36	ISATKUAG LAGOON DAM	BARROW	1977
## 37	ITASIGROOK	BARROW	1964
## 39	DEWEY FOREBAY	SKAGWAY	1905
## 40	LOWELL CREEK	SEWARD	1945
## 41	SHERWOOD ESTATES DAM	RAINBOW LAKE SUBDIVISION	1975
## 42	KOTZEBUE WATER SUPPLY DAM	KOTZEBUE	1970
## 47	DOUGLAS ISLAND DAM	DOUGLAS	1937
## 48	ROYCROFT LAKE	MOOSE PASS	1979
## 49	MOOSE CREEK DAM	FAIRBANKS	1979
## 50	MOOSE CREEK DAM - EAST CUTOFF DIKE	FAIRBANKS	1979
## 52	ALITAK CANNERY DAM #3	ALITAK CANNERY	1964
## 53	ALITAK CANNERY DAM #2	ALITAK CANNERY	1964
## 54	ALITAK CANNERY DAM #1	ALITAK CANNERY	1917
## 56	ALITAK CANNERY DAM #4	ALITAK CANNERY	1964
## 57	STOVER DAM	KODIAK	1962
## 58	BRIDGE CREEK DAM	HOMER	1975
## 59	CHIGNIK	CHIGNIK	1947
## 60	NORTH LAKE WEST DAM	ADAK NAVAL BASE	1945
## 61	NORTH LAKE EAST DAM	ADAK NAVAL BASE	1970
## 62	BONNIE ROSE LAKE DAM	ADAK NAVAL BASE	1943
## 63	LAKE LEONE DAM	ADAK NAVAL BASE	1945
## 64	ADAK LOG DAM	ADAK NAVAL BASE	1970
## 66	HYDABURG DAM	HYDABURG	1939

## 67	NEW ENGLAND FISH CO DAM	SAND POINT	1940
## 68	LAKE DEMARIE DAM	ADAK	1951
## 70	WARD COVE CANNERY DAM	KETCHIKAN	1932
## 74	LOWER FIRE LAKE DAM	EAGLE RIVER	1960
## 75	RED DOG WATER SUPPLY DAM	KIVALINA	1989
## 76	RED DOG TAILINGS MAIN DAM	KIVALINA	1993
## 77	MAHOONA DAM	OUZINKIE	1986
## 78	FORT KNOX WATER DAM	CHATINEKA	1995
## 79	FORT KNOX TAILINGS DAM	CHATINEKA	1996
## 80	NIXON FORK TAILINGS DAM	MCGRATH	1995
## 81	CABIN CREEK DAM	PETERSBURG	1996
## 82	RED DOG MINE WATER DIVERSION DAM	KIVALINA	1991
## 84	AKUTAN HYDRO DAM	AKUTAN	1993
## 85	ICY CREEK RESERVOIR DAM	CAPTAINS BAY	1993
## 86	ICY LAKE DAM	CAPTAINS BAY	1995
## 88	TWO BULL RIDGE POND T5 DAM	HEALY	2000
## 89	RED DOG TAILINGS BACK DAM	KIVALINA	2010
## 91	NEW KAKE DAM	KAKE	2007
## 96	A F K HATCHERY DAM	SAN JUAN CANNERY	1977
## 104	EXPLORER GLACIER POND DAM	PORTAGE	1989
##	MAX_STORAGE		
## 1	127600.0		
## 3	18000.0		
## 6	13353.0		
## 11	700.0		
## 13	190.0		
## 14	102.0		
## 15	142.0		
## 16	10200.0		
## 17	65.0		
## 18	45.0		
## 19	123.0		
## 20	93.0		
## 21	15.5		
## 24	800.0		
## 25	300.0		
## 27	58.0		
## 28	240000.0		
## 29	50.0		
## 30	27.0		
## 34	147000.0		
## 35	13000.0		
## 36	1140.0		
## 37	1300.0		
## 39	35.0		
## 40	50.0		
## 41	200.0		
## 42	412.0		
## 47	3.1		
## 48	3.6		
## 49	224000.0		
## 50	224000.0		
## 52	51.0		
## 53	52.0		

```
## 54      62.0
## 56      51.0
## 57     110.0
## 58     943.0
## 59     120.0
## 60      56.0
## 61      56.0
## 62    2260.0
## 63    1690.0
## 64     118.0
## 66       8.0
## 67     105.0
## 68    2120.0
## 70      46.0
## 74     550.0
## 75     684.0
## 76   40030.0
## 77     400.0
## 78    5350.0
## 79  136000.0
## 80      77.0
## 81    1270.0
## 82      58.0
## 84      20.0
## 85      30.0
## 86     238.0
## 88      20.9
## 89   24757.0
## 91      60.0
## 96     533.0
## 104     125.0
```

Analyze data

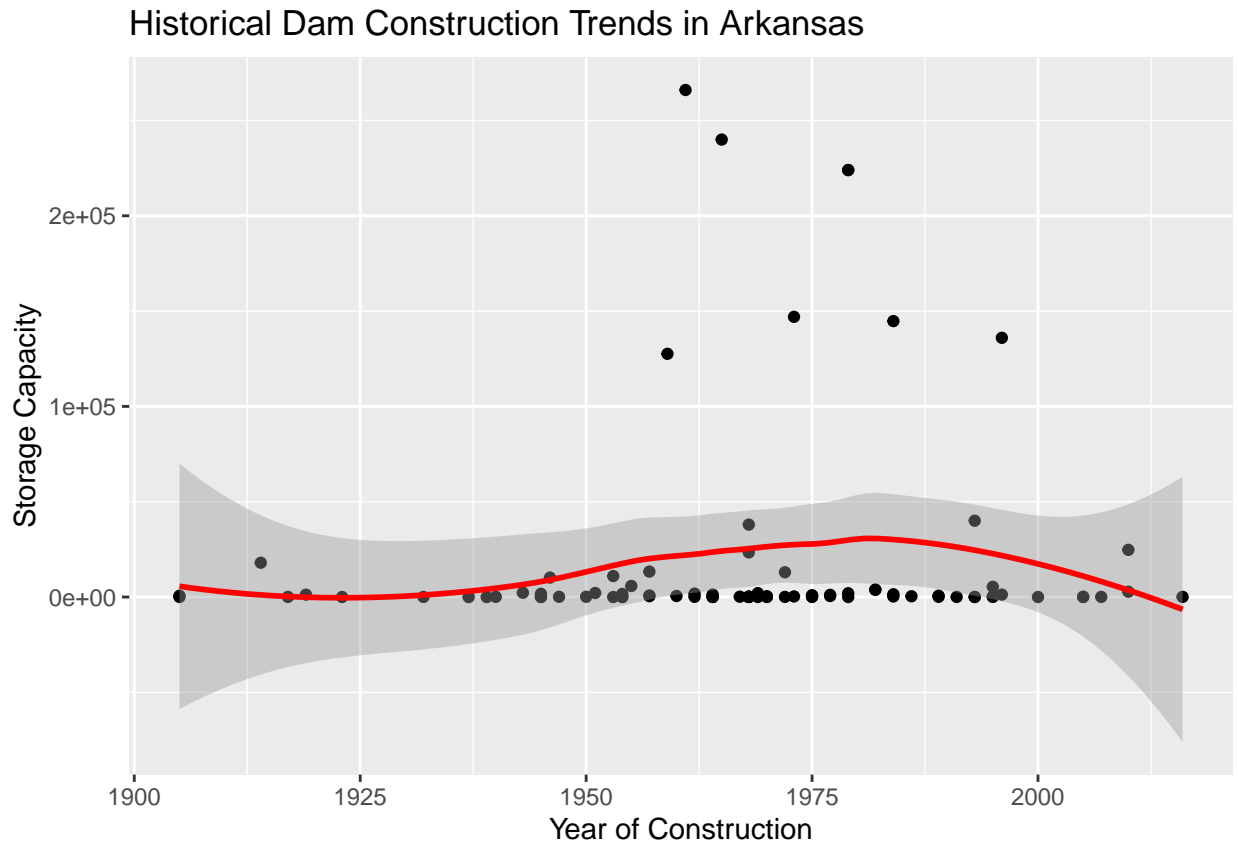
There is a gradual slight increase over time in dam storage capacity until about 1990 and then the capacity of new dams under construction generally decreases. Seven dams were had storage capacities that far exceeded the average dam capacity.

```
ggplot(dams)+
  geom_point(mapping = aes(dams$YEAR_COMPLETED, dams$MAX_STORAGE), show.legend=FALSE)+
  geom_smooth(mapping = aes(dams$YEAR_COMPLETED, dams$MAX_STORAGE), show.legend=FALSE, color = "red")+
  ggtitle("Historical Dam Construction Trends in Arkansas")+
  xlab("Year of Construction")+
  ylab("Storage Capacity")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

```
## Warning: Removed 17 rows containing non-finite values (stat_smooth).
```

```
## Warning: Removed 17 rows containing missing values (geom_point).
```



DATA SET 2: NYC AP Test Rates

<https://data.cityofnewyork.us/Education/2010-AP-College-Board-School-Level-Results/itfs-ms3e>

```
tests <- read.csv("https://raw.githubusercontent.com/jambawilliams/DATA607_PROJECT2/master/2010__AP__Co")
head(tests)
```

##	DBN	SchoolName	AP.Test.Takers	Total.Exams.Taken
## 1	01M448	UNIVERSITY NEIGHBORHOOD H.S.	39	49
## 2	01M450	EAST SIDE COMMUNITY HS	19	21
## 3	01M515	LOWER EASTSIDE PREP	24	26
## 4	01M539	NEW EXPLORATIONS SCI,TECH,MATH	255	377
## 5	02M296	High School of Hospitality Management	NA	NA
## 6	02M298	Pace High School	21	21
##	Number.of.Exams.with.scores.3.4.or.5			
## 1		10		
## 2		NA		
## 3		24		
## 4		191		
## 5		NA		
## 6		NA		

Clean data

Rename columns, null values, and empty cells

```

tests <- select(tests, -c("DBN"))
tests <- tests %>% rename( "School" = "SchoolName", "Taker_Count" = "AP.Test.Takers", "Exam_Count"="Total.Exam.Count")
tests %>%
  mutate_all(~ifelse(. %in% c("N/A", "NONE", "null", ""), NA, .)) %>%
  na.omit()%>%
  mutate(School = toupper(School))

```

##	School	Taker_Count	Exam_Count
## 1	UNIVERSITY NEIGHBORHOOD H.S.	39	49
## 2	LOWER EASTSIDE PREP	24	26
## 3	NEW EXPLORATIONS SCI,TECH,MATH	255	377
## 4	URBAN ASSEMBLY SCHOOL OF DESIGN AND CONSTRUCTION,	99	117
## 5	URBAN ASSEMBLY ACADEMY OF GOVERNMENT AND LAW, THE	25	37
## 6	HS FOR ENVIRONMENTAL STUDIES	213	298
## 7	PROFESSIONAL PERFORMING ARTS	20	20
## 8	BARUCH COLLEGE CAMPUS HS	78	115
## 9	NYC LAB HS FOR COLL. STUDIES	114	140
## 10	ELEANOR ROOSEVELT HIGH SCHOOL	155	235
## 11	MILLENNIUM HIGH SCHOOL	86	95
## 12	HEALTH PROF & HUMAN SVCS	204	248
## 13	H S FOR LEADERSHIP & PUBLIC	70	84
## 14	MANHATTAN VILLAGE ACADEMY HS	42	69
## 15	CES-MANHATTAN INTERNATIONAL	12	12
## 16	WASHINGTON IRVING HS	69	83
## 17	STUYVESANT HS	1510	2819
## 18	HS OF ECONOMICS & FINANCE	101	176
## 19	TALENT UNLIMITED	30	34
## 20	MURRY BERGTRAUM H.S.	118	157
## 21	JACQUELINE K. ONASSIS HS	80	118
## 22	MANHATTAN BRIDGES HIGH SCHOOL	67	102
## 23	HIGH SCHOOL FOR DUAL LANGUAGE AND ASIAN STUDIES	61	96
## 24	MANHATTAN COMP.NIGHT H.S.	58	100
## 25	R.R.GREEN HS OF TEACHING	67	117
## 26	HS FOR FASHION INDUSTRIES	70	87
## 27	CHELSEA VHS	52	97
## 28	NORMAN THOMAS H.S.	34	34
## 29	HS COMMUNICATION GRAPHIC ART	79	85
## 30	HS OF ART AND DESIGN	84	100
## 31	LIFE SCIENCES SECONDARY SCHL	50	90
## 32	LOUIS D. BRANDEIS HS	48	70
## 33	BEACON SCHOOL	166	197
## 34	FIORELLO H.LAGUARDIA HS	691	1236
## 35	MLK, JR HS FOR ARTS & TECH	38	38
## 36	MANHATTAN CENTER-SCIENCE&MATH	328	426
## 37	THE YOUNG WOMEN'S LEADERSHIP SCHOOL OF EAST HARLEM	41	55
## 38	FREDRICK DOUGLAS SEC. SCHOOL	163	226
## 39	HIGH SCHOOL FOR MATH SCIENCE ENGINEERING @ CCNY	114	124
## 40	H.S. INT'L BUSINESS & FINANCE	108	135
## 41	H.S. LAW & PUBLIC SERVICE	141	212
## 42	H.S. HEALTH CAREERS & SCIENCE	32	38
## 43	A PHILIP RANDOLPH CAMPUS	165	222
## 44	GREGORIO LUPERON HS SCI & MATH	88	138
## 45	SOUTH BRONX PREPARATORY: A COLLEGE BOARD SCHOOL	63	89

## 46	LINCOLN ACADEMY/ HOSTOS	64	107
## 47	FOREIGN LANG ACAD/GLOBAL STUD	13	13
## 48	BRONX ACADEMY OF LETTERS	53	92
## 49	HEALTH OPPORTUNITIES PROGRAM	61	117
## 50	RENAISSANCE HIGH SCHOOL FOR MUSICAL THEATER & TECH	22	29
## 51	HERBERT LEHMAN HS	430	532
## 52	JANE ADDAMS VHS	44	69
## 53	EAGLE ACADEMY FOR YOUNG MEN	32	44
## 54	HARRY S TRUMAN HS	103	135
## 55	EXIMIUS COLLEGE PREPARATORY ACADEMY: A COLLEGE BOA	90	146
## 56	BRONX CENTER FOR SCIENCE AND MATHEMATICS	62	80
## 57	LEADERSHIP INSTITUTE	17	17
## 58	ACADEMY FOR LANGUAGE AND TECHNOLOGY	20	20
## 59	BRONX HIGH SCHOOL OF BUSINESS	26	31
## 60	BRONX SCH FOR LAW, GOVT, JUST	45	55
## 61	BRONX LEADERSHIP ACADEMY HS	40	42
## 62	MS/HS 141 D A STEIN RIV/KINGSB	133	184
## 63	BRONX ENGINEERING AND TECHNOLOGY ACADEMY	107	228
## 64	WEST BRONX ACADEMY FOR THE FUTURE	25	29
## 65	INTERNATIONAL SCHOOL FOR LIBERAL ARTS	55	73
## 66	MS/HS 368 INFO & NETWORK TECH	40	54
## 67	HS FOR TEACHING & PROFESSIONS	62	83
## 68	BELMONT PREPARATORY HIGH SCHL	31	45
## 69	FORDHAM LDRSHP ACAD BUS & TECH	42	47
## 70	BRONX HS OF LAW & COMM SERVICE	44	65
## 71	DEWITT CLINTON HS	366	540
## 72	CELIA CRUZ BRONX HIGH SCHOOL OF MUSIC, THE	31	35
## 73	BRONX HS OF SCIENCE	1190	2435
## 74	JOHN F KENNEDY HS	46	48
## 75	MARBLE HILL HS FOR INT STUDIES	83	141
## 76	HIGH SCHOOL OF AMERICAN STUDIES AT LEHMAN COLLEGE	194	302
## 77	COLLEGIATE INSTITUTE FOR MATH AND SCIENCE	79	111
## 78	ASTOR COLLEGIATE ACADEMY	36	48
## 79	CHRISTOPHER COLUMBUS HS	34	34
## 80	BRONX HS FOR THE VISUAL ARTS	26	41
## 81	PELHAM PREPARATORY ACADEMY	46	57
## 82	BROOKLYN TECHNICAL HS	2117	3692
## 83	URBAN ASSEMBLY SCHOOL FOR LAW AND JUSTICE, THE	31	57
## 84	BEDFORD ACADEMY HIGH SCHOOL	43	57
## 85	BENJAMIN BANNEKER ACADEMY	132	149
## 86	JHS 071 JUAN MOREL CAMPOS	16	16
## 87	HS OF ENTERPRISE AND TECH	94	129
## 88	BROOKLYN PREPARATORY HIGH SCHOOL	28	46
## 89	ACADEMY FOR COLLEGE PREPARATION AND CAREER EXPLORA	64	82
## 90	HIGH SCHOOL FOR PUBLIC SERVICE: HEROES OF TOMORRO	106	130
## 91	BROOKLYN SCHOOL FOR MUSIC & THEATRE	37	41
## 92	MIDDLE COLL HS @ MEDGAR EVERS	272	507
## 93	CLARA BARTON VHS	135	211
## 94	MULTICULTURAL HIGH SCHOOL	44	44
## 95	ENY-TRANSIT TECH HS	138	164
## 96	CYPRESS HILLS COLLEGIATE PREPARATORY SCHOOL	21	21
## 97	NEW UTRECHT HS	174	271
## 98	HS OF TELECOMMUNICATIONS	251	413
## 99	FORT HAMILTON HS	357	576

## 100	F.D. ROOSEVELT HS	228	340
## 101	ABRAHAM LINCOLN HS	133	193
## 102	EDWARD R MURROW HS	397	577
## 103	JOHN DEWEY H.S.	232	370
## 104	BROOKLYN STUDIO SCHOOL	39	43
## 105	MIDWOOD HS	745	1223
## 106	JAMES MADISON HS	323	463
## 107	SHEEPSHEAD BAY HS	37	59
## 108	LEON GOLDSTEIN HS SCIENCES	367	575
## 109	ACADEMY OF FINANCE AND ENTERPRISE	95	137
## 110	NEWTOWN HS	174	200
## 111	GROVER CLEVELAND HS	200	274
## 112	HIGH SCHOOL FOR ARTS & BUSINESS	198	320
## 113	QUEENS VHS	27	27
## 114	AVIATION CAREER & TECHNICAL H.S.	130	140
## 115	WORLD JOURNALISM PREPARATORY: A COLLEGE BOARD SCHO	65	87
## 116	JOHN BOWNE HS	176	229
## 117	FLUSHING HS	324	424
## 118	TOWNSEND HARRIS HS	613	796
## 119	R F KENNEDY COLLABORATIVE H S	77	112
## 120	BENJAMIN N. CARDOZO	676	1145
## 121	FRANCIS LEWIS HS	697	1033
## 122	MARTIN VAN BUREN HS	71	101
## 123	BAYSIDE HS	480	686
## 124	CHANNEL VIEW SCHOOL FOR RESEARCH	43	69
## 125	SCHOLARS' ACADEMY	55	121
## 126	RICHMOND HILL HS	142	170
## 127	JOHN ADAMS HS	140	168
## 128	HIGH SCHOOL FOR CONSTRUCTION TRADES, ENGINEERING A	91	109
## 129	FOREST HILLS HS	733	1116
## 130	JAMAICA HS	64	76
## 131	HILLCREST HS	241	362
## 132	THOS A EDISON VHS	183	244
## 133	GATEWAY TO HEALTH SCIENCES	89	136
## 134	QUEENS HS FOR SCIENCE YORK COLLEGE	215	338
## 135	MATH/SCIENCE RESEARCH TECH CTR	22	31
## 136	WM CULLEN BRYANT HS	243	330
## 137	LONG ISLAND CITY HS	406	805
## 138	FRANK SINATRA HIGH SCHOOL	108	136
## 139	INFORMATION TECHNOLOGY HS	51	74
## 140	NEWCOMERS HS:ACAD AMER STD	119	163
## 141	ACADEMY OF AMER. STUDIES HS	101	170
## 142	PS 080 MICHAEL J. PETRIDES	162	243
## 143	NEW DORP HS	96	118
## 144	PORT RICHMOND HS	194	304
## 145	CURTIS HS	247	295
## 146	TOTTENVILLE HS	396	687
## 147	SUSAN E. WAGNER HS	279	408
## 148	STATEN ISLAND TECHNICAL HS	528	905
## 149	EBC-HS FOR PUB SERVICE (BUSH)	47	64
## 150	ACADEMY OF URBAN PLANNING	76	100
## 151	BUSHWICK LEADERS HIGH SCHOOL FOR ACADEMIC EXCELLEN	34	35
##	Pass_Count		
## 1	10		

## 2	24
## 3	191
## 4	10
## 5	15
## 6	152
## 7	15
## 8	88
## 9	97
## 10	169
## 11	67
## 12	75
## 13	32
## 14	29
## 15	6
## 16	21
## 17	2648
## 18	78
## 19	21
## 20	76
## 21	29
## 22	59
## 23	89
## 24	61
## 25	40
## 26	26
## 27	13
## 28	14
## 29	9
## 30	33
## 31	10
## 32	6
## 33	140
## 34	790
## 35	6
## 36	261
## 37	29
## 38	159
## 39	71
## 40	43
## 41	82
## 42	17
## 43	72
## 44	73
## 45	24
## 46	39
## 47	12
## 48	37
## 49	29
## 50	14
## 51	197
## 52	8
## 53	11
## 54	27
## 55	11

## 56	35
## 57	7
## 58	20
## 59	9
## 60	27
## 61	8
## 62	119
## 63	10
## 64	14
## 65	45
## 66	17
## 67	13
## 68	10
## 69	12
## 70	9
## 71	262
## 72	8
## 73	2189
## 74	19
## 75	37
## 76	243
## 77	33
## 78	12
## 79	6
## 80	9
## 81	17
## 82	2687
## 83	24
## 84	16
## 85	62
## 86	12
## 87	19
## 88	7
## 89	7
## 90	13
## 91	9
## 92	86
## 93	63
## 94	39
## 95	18
## 96	10
## 97	95
## 98	123
## 99	331
## 100	186
## 101	18
## 102	307
## 103	154
## 104	13
## 105	758
## 106	188
## 107	27
## 108	382
## 109	40

## 110	117
## 111	112
## 112	73
## 113	20
## 114	61
## 115	18
## 116	116
## 117	136
## 118	625
## 119	43
## 120	796
## 121	583
## 122	15
## 123	413
## 124	18
## 125	57
## 126	55
## 127	33
## 128	21
## 129	526
## 130	8
## 131	50
## 132	100
## 133	57
## 134	275
## 135	8
## 136	97
## 137	169
## 138	38
## 139	13
## 140	140
## 141	119
## 142	66
## 143	62
## 144	76
## 145	90
## 146	206
## 147	175
## 148	809
## 149	13
## 150	10
## 151	18

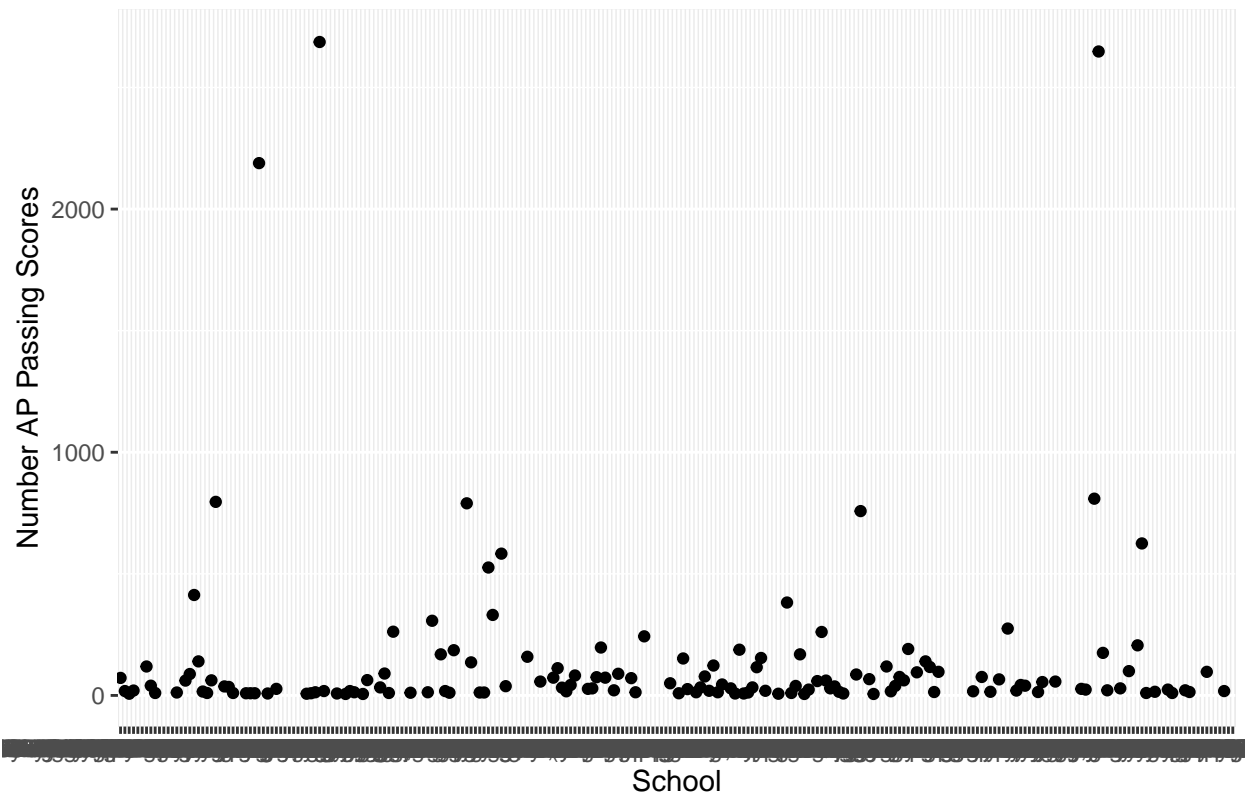
Analyze data

There are about a dozen schools in NYC that have AP pass counts far above average counts. This could be due to larger study body and AP student enrollment among other factors.

```
ggplot(tests)+
  geom_point(mapping = aes(tests$School, tests$Pass_Count), show.legend=FALSE)+
  ggtitle("NYC Schools 2010 AP Pass Counts")+
  xlab("School")+
  ylab("Number AP Passing Scores")
```

```
## Warning: Removed 107 rows containing missing values (geom_point).
```

NYC Schools 2010 AP Pass Counts



DATA SET 3: NYC Gifted and Talented Scores

<https://makingnoiseandhearingthings.com/2018/04/19/datasets-for-data-cleaning-practice/>

```
gifted <- read.csv("https://raw.githubusercontent.com/jambawilliams/DATA607_PROJECT2/master/G%26T%20Res")
head(gifted)
```

```
##   Timestamp Entering.Grade.Level District Birth.Month OLSAT.Verbal.Score
## 1  3/27/18                K         2           4             29
## 2  3/28/18                K         2           4             25
## 3  3/29/18                K         2           8             30
## 4  3/30/18                K         2           8             30
## 5  3/31/18                K         2           1             25
## 6  4/1/18                 K         6           7             25
##   OLSAT.Verbal.Percentile NNAT.Non.Verbal.Raw.Score NNAT.Non.Verbal.Percentile
## 1                      99                      37                      99
## 2                      99                      38                      99
## 3                      99                      46                    Illinois
## 4                      99                      46                    Illinois
## 5                      99                      45                    Illinois
## 6                      99                      42                    Illinois
##   Overall.Score School.Assigned X Will.you.enroll.there.
## 1           99             NA
```

```
## 2          99          NA at LL now, has a sib ent K w/ 99
## 3          99          Anderson? NA
## 4          99          Anderson? NA          Yes
## 5          99          NEST+M NA
## 6          99 Anderson, TAG, Nest NA
## Test.Prep.
## 1          Y
## 2          No
## 3          No
## 4          at home
## 5
## 6          Y
```

Clean data

Rename columns, null values, and empty cells

```
gifted <- gifted[, c("District", "Overall.Score")]
gifted <- gifted %>% rename("Score" = "Overall.Score")
gifted %>%
  mutate_all(~ifelse(. %in% c("N/A", "NONE", "null", ""), NA, .)) %>%
  na.omit()
```

```
## District Score
## 1      2    99
## 2      2    99
## 3      2    99
## 4      2    99
## 5      2    99
## 6      6    99
## 7     30    99
## 8      2    99
## 9      2    99
## 10     13    99
## 11     21    99
## 13      2    99
## 14     30    99
## 15      2    99
## 16      2    97
## 17     27    96
## 18      2    99
## 19      3    99
## 20      2    99
## 21      2    99
## 22      2    99
## 23     28    99
```

Analyze data

Only eight districts have gifted student placements. The vast majority of gifted student are in District 2.

```
ggplot(gifted, aes(District)) +  
  geom_bar() +  
  ggtitle("NYC Gifted Students Placement")+  
  xlab("School District")+  
  ylab("Count of Gifted Students")
```

Warning: Removed 1 rows containing non-finite values (stat_count).

