

R을 이용한 범죄발생 빅데이터 분석 및 시각화

ggplot2 지도, 막대그래프, 원형 그래프 / Pie Chart 시각화

지도교수: 김선옥

201632066 전우진

목차

- 01. 연구 주제 선정 및 필요성
- 02. 연구 진행 계획
- 03. 다양한 데이터를 이용한 시각화 학습
- 04. 연구 과정 및 결과
- 05. 시사점 및 한계점
- 06. Q&A

01. 연구 주제 선정 및 필요성



01. 연구 주제 선정 및 필요성

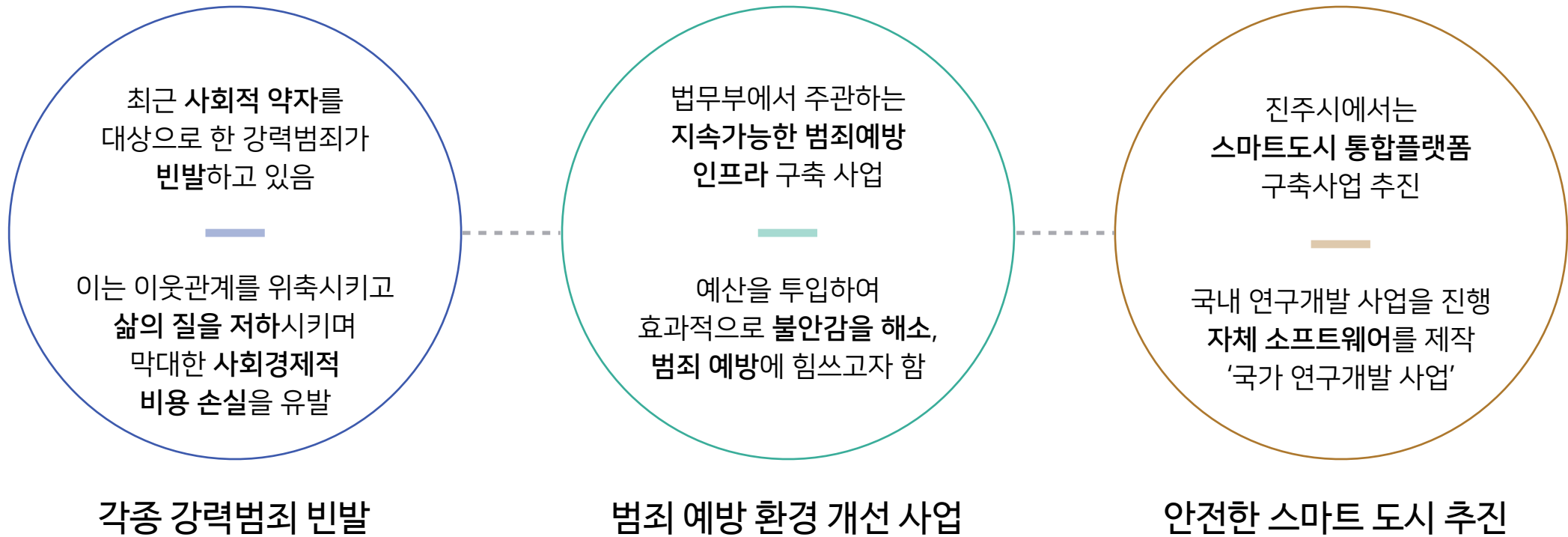
'R을 이용한 범죄발생 빅데이터 분석 및 시각화'

연구주제 | R 프로그램을 이용한 범죄 데이터 분석 및 시각화

필요성 | 사람들의 **웰빙 욕구**가 증가하면서 동시에 '**안전한 도시**'에 대한 **관심도** 또한 높아지고 있다. 따라서 안전한 도시를 만들기 위한 리빙랩 프로젝트, 사회 안전망 구축 등 **다양한 연구가** 진행되고 있다. 연구 진행 시에 범죄 발생률 등 관련 데이터를 분석하는 것은 필수적인 요소로서, 사회 불안감을 해소하기 위해 **데이터를 수집하고 분석하는 과정**이 중요하게 대두되고 있다.



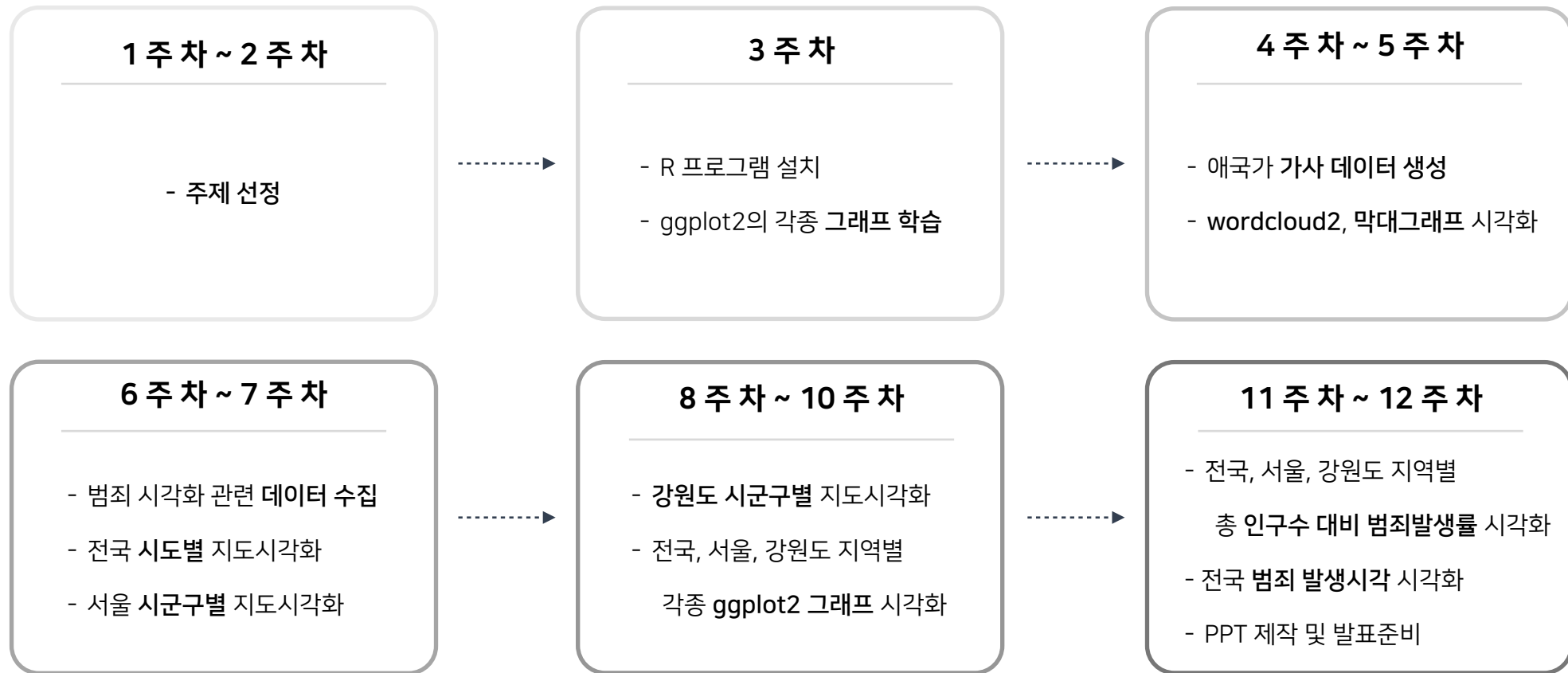
01. 연구 주제 선정 및 필요성



02. 연구 진행 계획



02. 연구 진행 계획



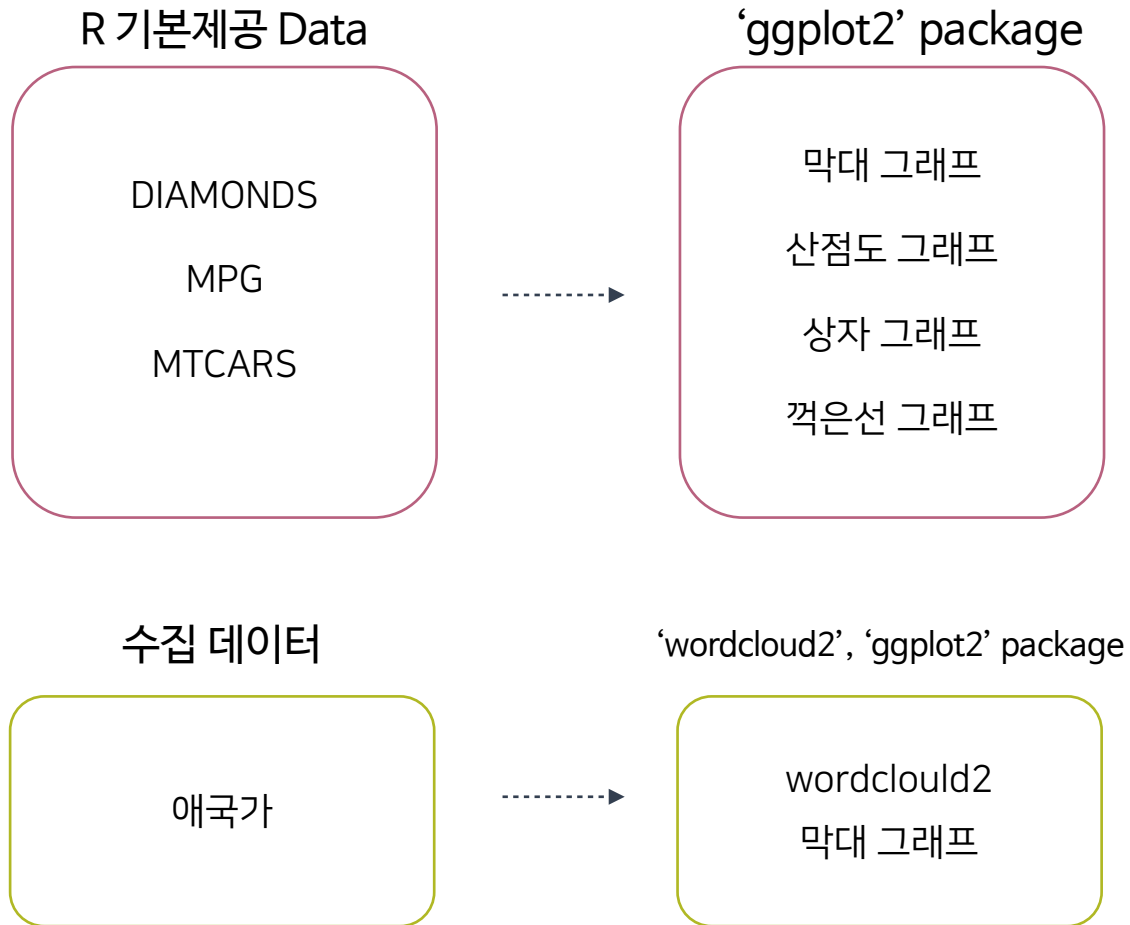
주제 선정 -----> 관련 사례 및 선행연구 조사 -----> R 시각화를 위한 학습 진행
-----> 학습한 내용 기반, 선정된 주제에 관련된 데이터 분석 진행 -----> 결과 도출

03. 다양한 데이터를 이용한 시각화 학습



03. 다양한 데이터를 이용한 시각화 학습

R 기본제공 데이터 분석 및 각종 'ggplot2' 그래프 시각화



[3주차 ~ 5주차]

- R 기본 제공 데이터와 수집 데이터를 분석 후 시각화 하였음
- ggplot2 패키지의 여러 그래프 시각화 방법을 학습



03. 다양한 데이터를 이용한 시각화 학습

1) DIAMONDS DATA ----> 'ggplot2' 막대그래프

	carat	cut	color	clarity	depth	table	price	x	y	z
1	0.23	Ideal	E	SI2	61.5	55.0	326	3.95	3.98	2.43
2	0.21	Premium	E	SI1	59.8	61.0	326	3.89	3.84	2.31
3	0.23	Good	E	VS1	56.9	65.0	327	4.05	4.07	2.31
4	0.29	Premium	I	VS2	62.4	58.0	334	4.20	4.23	2.63
5	0.31	Good	J	SI2	63.3	58.0	335	4.34	4.35	2.75
6	0.24	Very Good	J	VVS2	62.8	57.0	336	3.94	3.96	2.48
7	0.24	Very Good	I	VVS1	62.3	57.0	336	3.95	3.98	2.47
8	0.26	Very Good	H	SI1	61.9	55.0	337	4.07	4.11	2.53
9	0.22	Fair	E	VS2	65.1	61.0	337	3.87	3.78	2.49
10	0.23	Very Good	H	VS1	59.4	61.0	338	4.00	4.05	2.39
11	0.30	Good	J	SI1	64.0	55.0	339	4.25	4.28	2.73
12	0.23	Ideal	J	VS1	62.8	56.0	340	3.93	3.90	2.46
13	0.22	Premium	F	SI1	60.4	61.0	342	3.88	3.84	2.33
14	0.31	Ideal	J	SI2	62.2	54.0	344	4.35	4.37	2.71
15	0.20	Premium	E	SI2	60.2	62.0	345	3.79	3.75	2.27
16	0.32	Premium	E	I1	60.9	58.0	345	4.38	4.42	2.68

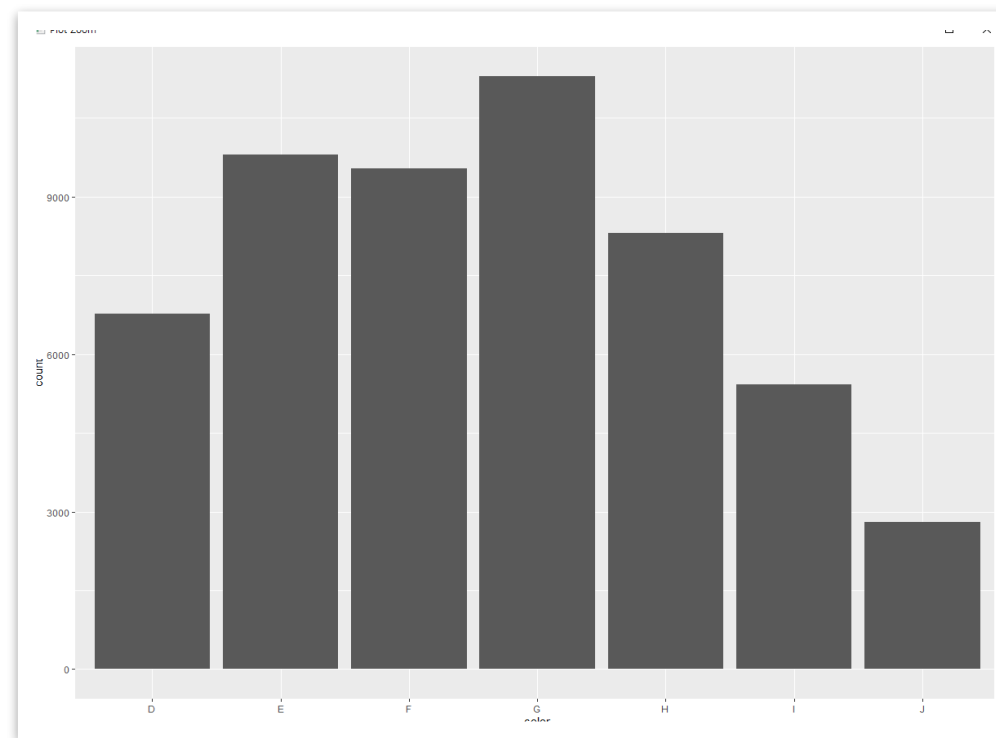
53928	0.79	Good	F	SI1	58.1	59.0	2756	6.06	6.13	3.54
53929	0.79	Premium	E	SI2	61.4	58.0	2756	6.03	5.96	3.68
53930	0.71	Ideal	G	VS1	61.4	56.0	2756	5.76	5.73	3.53
53931	0.71	Premium	E	SI1	60.5	55.0	2756	5.79	5.74	3.49
53932	0.71	Premium	F	SI1	59.8	62.0	2756	5.74	5.73	3.43
53933	0.70	Very Good	E	VS2	60.5	59.0	2757	5.71	5.76	3.47
53934	0.70	Very Good	E	VS2	61.2	59.0	2757	5.69	5.72	3.49
53935	0.72	Premium	D	SI1	62.7	59.0	2757	5.69	5.73	3.58
53936	0.72	Ideal	D	SI1	60.8	57.0	2757	5.75	5.76	3.50
53937	0.72	Good	D	SI1	63.1	55.0	2757	5.69	5.75	3.61
53938	0.70	Very Good	D	SI1	62.8	60.0	2757	5.66	5.68	3.56
53939	0.86	Premium	H	SI2	61.0	58.0	2757	6.15	6.12	3.74

[사진1] R 제공 RAW 데이터

54,000여 개의 다이아몬드에 대한 가격을 포함한 10개의 속성 정보

분석 과정

- R에서 기본으로 제공되는 **diamonds데이터**를 이용하여 x, y축의 특정 항목을 지정.
- ggplot2 패키지를 사용하여 **막대그래프**로 시각화 함.



[사진2] ggplot2 막대 그래프 시각화

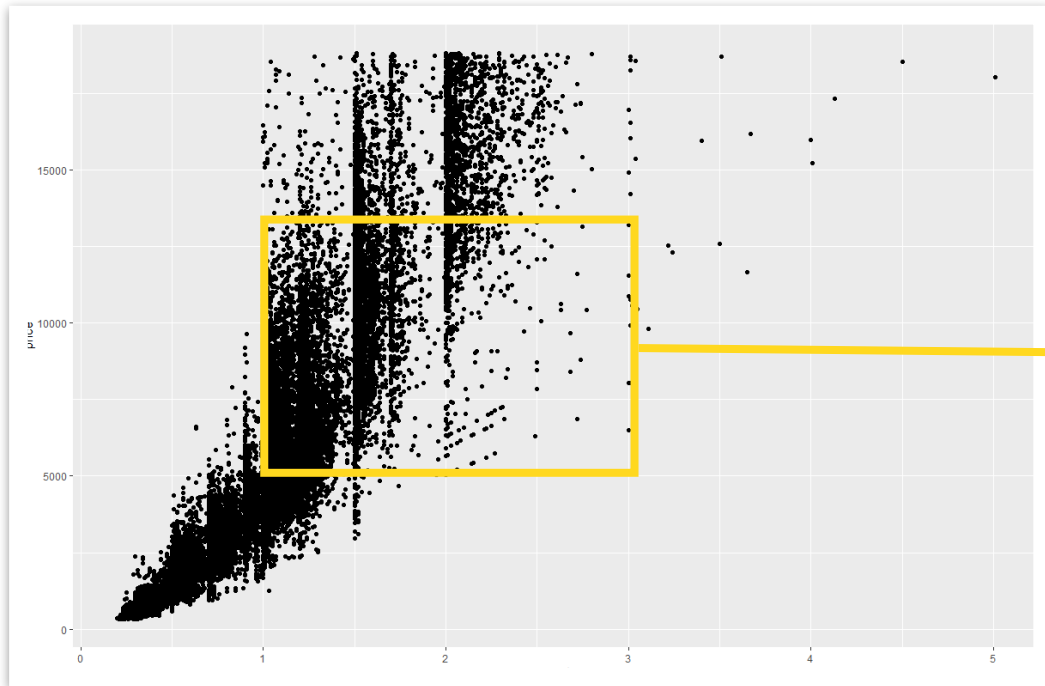
x축: 다이아몬드 색상 y축: 데이터 정보수량



03. 다양한 데이터를 이용한 시각화 학습

2) DIAMONDS DATA → 'ggplot2' 산점도 그래프

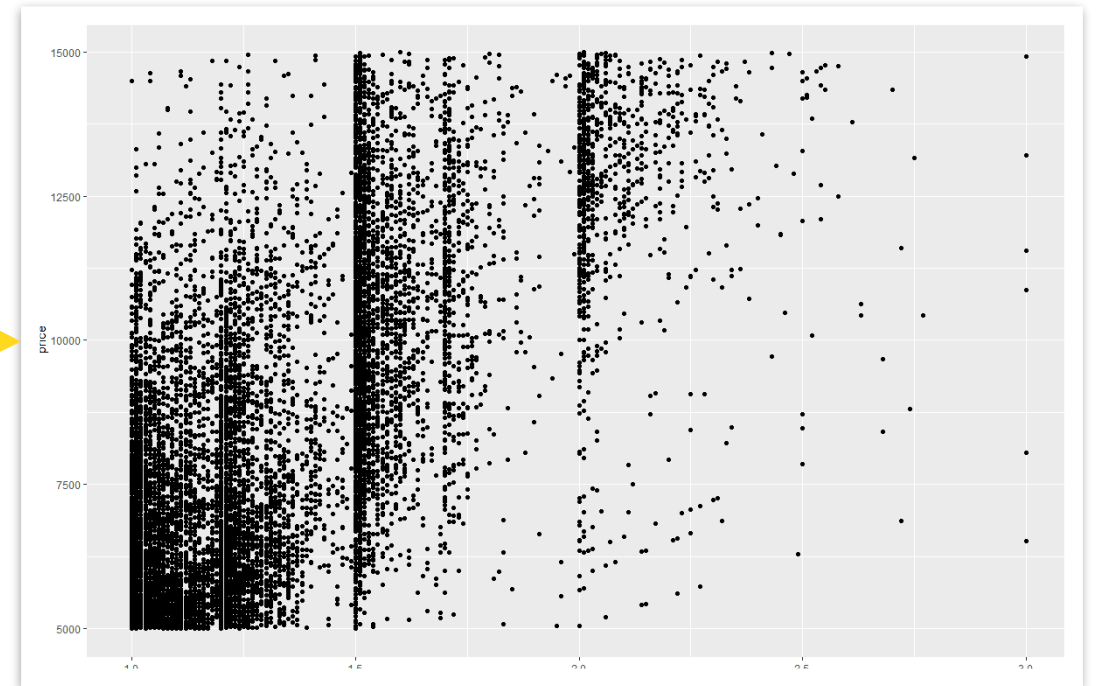
- 분석 과정
- diamonds데이터를 이용하여 x, y축의 특정 항목을 지정
 - ggplot2 패키지를 사용하여 산점도 그래프로 시각화



[사진3] ggplot2 산점도 그래프 시각화

x축: Carat(캐럿) y축: Price(가격)

- 분석 과정
- x, y축의 범위 지정 후 시각화



[사진4] x, y축 특정 범위 지정 후 시각화한 ggplot2 산점도 그래프



03. 다양한 데이터를 이용한 시각화 학습

3) MPG DATA ---▶ 'ggplot2' 상자 그래프

	manufacturer	model	displ	year	cyl	trans	drv	cty	hwy	fl	class
1	audi	a4	1.8	1999	4	auto(l5)	f	18	29	p	compact
2	audi	a4	1.8	1999	4	manual(m5)	f	21	29	p	compact
3	audi	a4	2.0	2008	4	manual(m6)	f	20	31	p	compact
4	audi	a4	2.0	2008	4	auto(av)	f	21	30	p	compact
5	audi	a4	2.8	1999	6	auto(l5)	f	16	26	p	compact
6	audi	a4	2.8	1999	6	manual(m5)	f	18	26	p	compact
7	audi	a4	3.1	2008	6	auto(av)	f	18	27	p	compact
8	audi	a4 quattro	1.8	1999	4	manual(m5)	4	18	26	p	compact
9	audi	a4 quattro	1.8	1999	4	auto(l5)	4	16	25	p	compact
10	audi	a4 quattro	2.0	2008	4	manual(m6)	4	20	28	p	compact
11	audi	a4 quattro	2.0	2008	4	auto(s6)	4	19	27	p	compact
12	audi	a4 quattro	2.8	1999	6	auto(l5)	4	15	25	p	compact
13	audi	a4 quattro	2.8	1999	6	manual(m5)	4	17	25	p	compact
14	audi	a4 quattro	3.1	2008	6	auto(s6)	4	17	25	p	compact
15	audi	a4 quattro	3.1	2008	6	manual(m6)	4	15	25	p	compact
16	audi	a6 quattro	2.8	1999	6	auto(l5)	4	15	24	p	midsize
17	audi	a6 quattro	3.1	2008	6	auto(s6)	4	17	25	p	midsize
18	audi	a6 quattro	4.2	2008	8	auto(s6)	4	16	23	p	midsize
19	chevrolet	c1500 suburban 2wd	5.3	2008	8	auto(l4)	r	14	20	r	suv

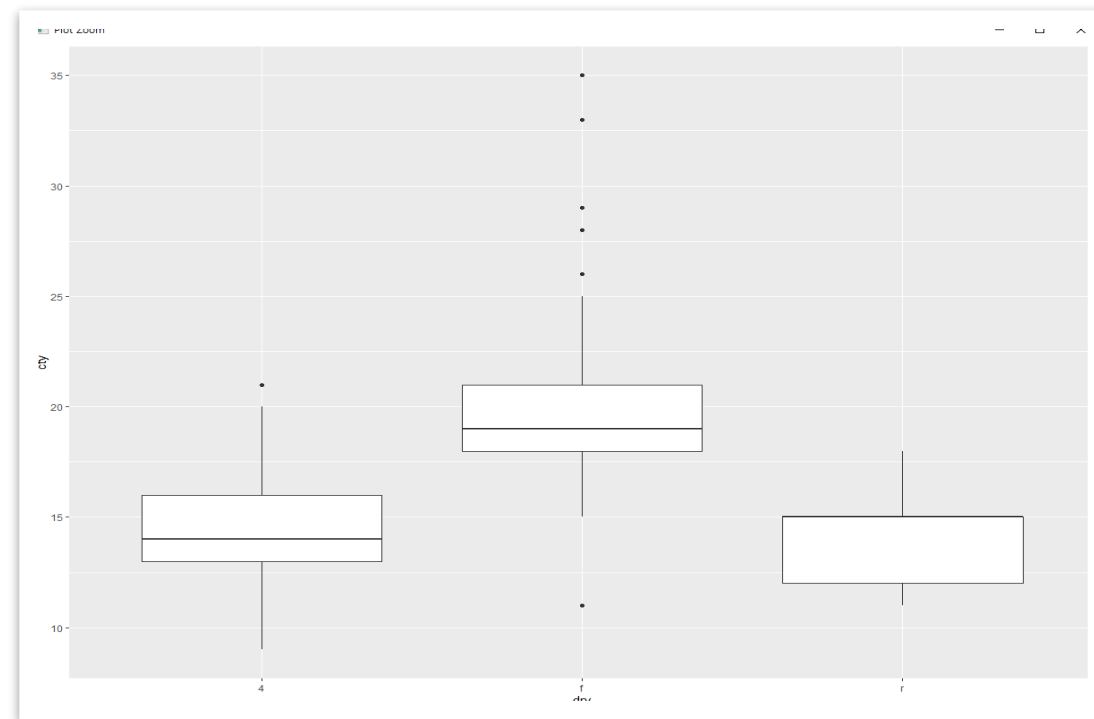
221	volkswagen	jetta	2.8	1999	6	manual(m5)	f	17	24	r	compact
222	volkswagen	new beetle	1.9	1999	4	manual(m5)	f	35	44	d	subcompact
223	volkswagen	new beetle	1.9	1999	4	auto(l4)	f	29	41	d	subcompact
224	volkswagen	new beetle	2.0	1999	4	manual(m5)	f	21	29	r	subcompact
225	volkswagen	new beetle	2.0	1999	4	auto(l4)	f	19	26	r	subcompact
226	volkswagen	new beetle	2.5	2008	5	manual(m5)	f	20	28	r	subcompact
227	volkswagen	new beetle	2.5	2008	5	auto(s6)	f	20	29	r	subcompact
228	volkswagen	passat	1.8	1999	4	manual(m5)	f	21	29	p	midsize
229	volkswagen	passat	1.8	1999	4	auto(l5)	f	18	29	p	midsize
230	volkswagen	passat	2.0	2008	4	auto(s6)	f	19	28	p	midsize
231	volkswagen	passat	2.0	2008	4	manual(m6)	f	21	29	p	midsize
232	volkswagen	passat	2.8	1999	6	auto(l5)	f	16	26	p	midsize
233	volkswagen	passat	2.8	1999	6	manual(m5)	f	18	26	p	midsize
234	volkswagen	passat	3.6	2008	6	auto(s6)	f	17	26	p	midsize

[사진5] R 제공 RAW 데이터

각종 차량의 배기량, 연비 데이터

분석 과정

- R에서 기본으로 제공되는 **mpg** 데이터를 이용하여 x, y축의 특정 항목을 지정
- ggplot2 패키지를 사용하여 **상자그래프**로 시각화



[사진6] ggplot2 상자 그래프 시각화

x축: 구동방식별 y축: 시내주행 연비



03. 다양한 데이터를 이용한 시각화 학습

4) MTCARS DATA ---> 'ggplot2' 꺾은선 그래프

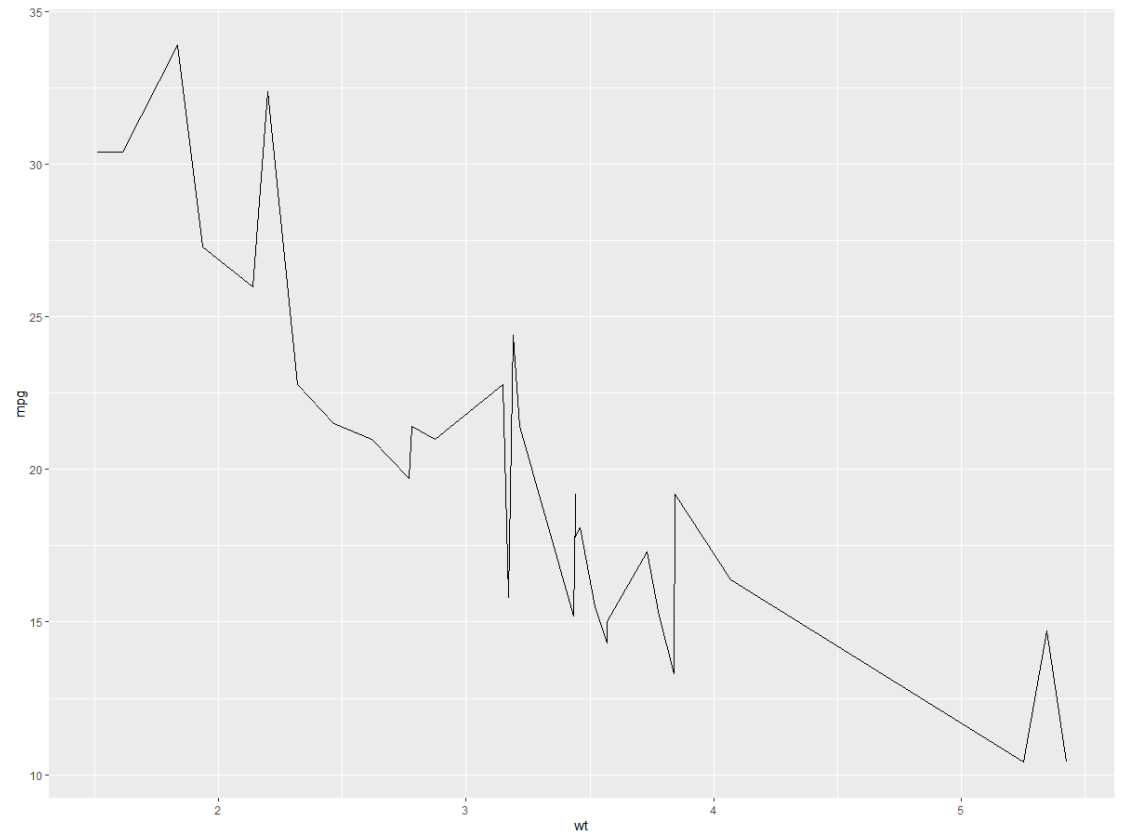
	mpg	cyl	displacement	horsepower	drat	weight	quarter mile time	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3
Cadillac Fleetwood	10.4	8	472.0	205	2.93	5.250	17.98	0	0	3	4
Lincoln Continental	10.4	8	460.0	215	3.00	5.424	17.62	0	0	3	4
Chrysler Imperial	14.7	8	440.0	230	3.23	5.345	17.42	0	0	3	4
Fiat 128	32.4	4	78.7	66	4.08	2.200	19.47	1	1	4	1
Honda Civic	30.4	4	75.7	52	4.93	1.615	18.52	1	1	4	2
Toyota Corolla	33.9	4	71.1	65	4.22	1.835	19.90	1	1	4	1
Toyota Corona	21.5	4	120.1	97	3.70	2.465	20.01	1	0	3	1
Dodge Challenger	15.5	8	318.0	150	2.76	3.520	16.87	0	0	3	2
AMC Javelin	15.2	8	304.0	150	3.15	3.435	17.30	0	0	3	2
Camaro Z28	13.3	8	350.0	245	3.73	3.840	15.41	0	0	3	4
Pontiac Firebird	19.2	8	400.0	175	3.08	3.845	17.05	0	0	3	2
Fiat X1-9	27.3	4	79.0	66	4.08	1.935	18.90	1	1	4	1
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.70	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.90	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.50	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.50	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.60	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.60	1	1	4	2

[사진7] R 제공 RAW 데이터

32종 자동차의 10가지 디자인과 성능특성과 연료소모량

분석 과정

- R에서 기본으로 제공되는 **mtcars** 데이터를 이용하여 x, y축의 특정 항목을 지정
- ggplot2 패키지를 사용하여 **꺾은선그래프**로 시각화



[사진8] ggplot 꺾은선 그래프 시각화

X축: 차량 중량 y축: mpg(연비)



03. 다양한 데이터를 이용한 시각화 학습

5) 애국가 가사 DATA → 'wordcloud2' 'ggplot2' 막대 그래프

a - Windows 메모장

파일(F) 편집(E) 서식(O) 보기(V) 도움말(H)

절
동해물과 백두산이 마르고 닳도록
하느님이 보우하사 우리나라만세
무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세

절
삼산위에 저 소나무 철갑을 두른듯
바람서리 불변함은 우리 기상일세
무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세

절
가을하늘 공활한데 높고 구름없이
바은달은 우리 가슴 일편단심일세
무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세

절
기 기상과 이 맘으로 충성을 다하여
괴로우나 즐거우나 나라사랑하세
무궁화 삼천리 화려강산 대한사람 대한으로 길이 보전하세

[사진9] 애국가 가사 데이터

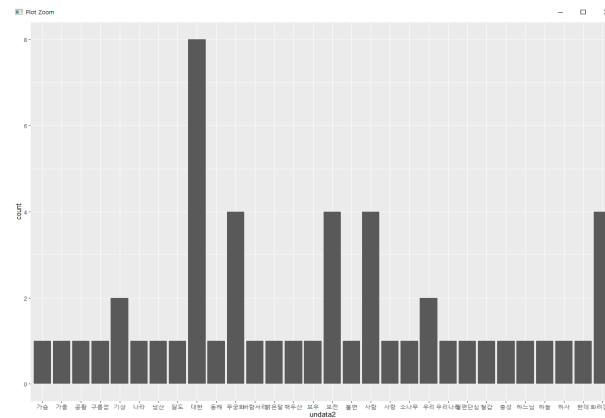
- 분석과정1 에서 '세종사전'을 사용하여 시각화하였을 때, '단어'의 단위로 추출되지 않고, 리, 하, 세 등 단어가 아닌 글자로 추출되는 **오류 발생**
- 여러 조건을 추가하여 **정확도 높임**

분석 과정 1

- R 내장 '세종사전'을 사용하여 데이터에서 **사전에 등록된 명사** 만을 추출
- 빈도수 값을 구하기 위해 데이터 타입 변경



[사진10] 키워드 wordcloud2 시각화



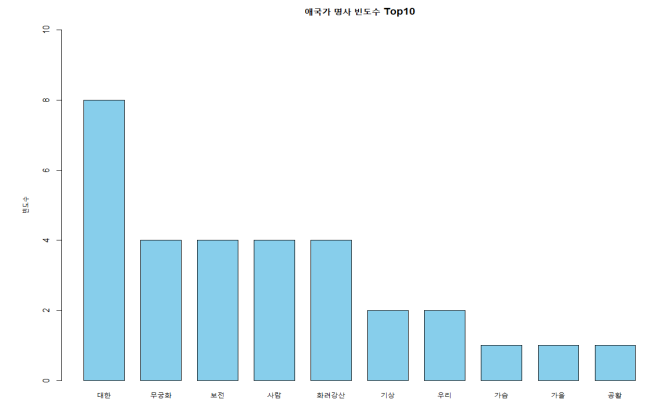
[사진11] 키워드별 빈도수 ggplot2 막대 그래프 시각화

분석 과정 2

- 직접 명사를 입력하여 사전에 등록
- 2글자 이상만 추출되도록 조건 설정



[사진12] 명사(형용사 부사 제외) wordcloud2 시각화



[사진13] 명사 빈도수 ggplot2 막대 그래프 시각화

04. 연구 과정 및 결과



04. 연구 과정 및 결과

범죄 관련 데이터 수집 후 다양하게 시각화

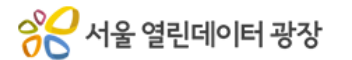
[분석 대상 데이터]

지역별 범죄 통계 데이터
지역별 코드(id) 데이터
지역별 좌표(위도, 경도) 데이터
지역별 GIS 행정구역 데이터
지역별 총 인구수
전국 범죄발생 시간

['ggplot2' package, Pie Chart]

- 지역별 범죄 발생 수
 - 지도 시각화
 - 막대 그래프
 - 원형 그래프
 - Pie Chart
- 지역별 인구수대비 범죄 발생률
 - 막대 그래프
- 전국 범죄 발생 시간
 - 원형 그래프

- 범죄 발생 현황 지도 시각화 관련 데이터 수집
- 시각화에 요구되는 데이터로 분석 및 변환
- 다양한 시각화 수행



[데이터 수집 기관 및 플랫폼]



04. 연구 과정 및 결과

- 분석 대상 데이터 수집 예시



[사진14] 키워드 GIS 행정구역 데이터

	A	B	C	D	E
1	시도	연령별	인구수	남자인구수	여자인구수
2	전국	계	51778544	25855919	25922625
3	전국	0세	345786	178138	167648
4	전국	1세	409814	209813	200001
5	전국	2세	442943	227063	215880
6	전국	3세	439700	225268	214432
7	전국	4세	440872	225898	214974
8	전국	5세	489401	251163	238238
9	전국	6세	476061	244330	231731
10	전국	7세	473580	244138	229442
11	전국	8세	448183	230769	217414
12	전국	9세	468858	241269	227589
13	전국	10세	496041	255068	240973
14	전국	11세	449954	232656	217298
15	전국	12세	436496	226040	210456
16	전국	13세	474366	245931	228435
17	전국	14세	493910	256513	237397
18	전국	15세	495262	258638	236624
19	전국	16세	559804	291386	268418
20	전국	17세	639416	334156	305260
21	전국	18세	610744	322172	288572

[사진16] 전국 시도별 인구수 데이터

	A	B	C
1	시도	long	lat
2	서울특별시	126.9899	37.551
3	부산광역시	129.0412	35.16336
4	대구광역시	128.5633	35.82527
5	인천광역시	126.7067	37.45599
6	광주광역시	126.8275	35.15438
7	대전광역시	127.3878	36.34908
8	울산광역시	129.2516	35.54521
9	세종특별자치시	127.2567	36.56095
10	경기도	127.3967	37.49523
11	강원도	128.163	37.80924
12	충청북도	127.7428	36.8694
13	충청남도	126.7925	36.60968
14	전라북도	127.1413	35.71474
15	전라남도	126.9655	34.86283
16	경상북도	128.8881	36.28358
17	경상남도	128.196	35.44669
18	제주특별자치도	126.5403	33.38957

[사진17] 전국 시도별 좌표 데이터

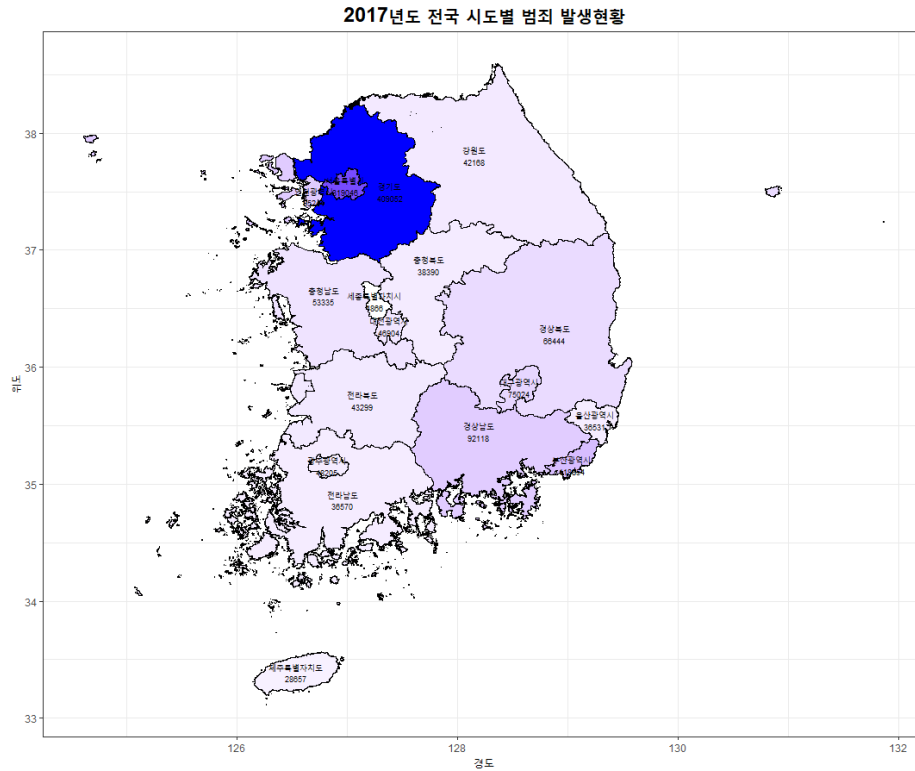
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	범죄대분류	범죄중분류	서울	부산	대구	인천	광주	대전	울산	세종	경기 고양	경기 과천	경기 광명	경기 광주	경기 구리	경기 군포	경기 김포	경기 남양주	경기 동두천	경기 부천	경기 성남	경기 수원	경기 시흥	경기 안산	경기 안성	경기 양주
2	강력범죄	살인미수	50	21	16	11	3	8	5	1	4	0	1	1	0	0	1	1	0	2	4	2	2	4	1	
3	강력범죄	살인미수	91	41	30	24	6	14	12	5	5	0	0	5	4	2	2	8	0	1	15	12	4	16	1	
4	강력범죄	강도	167	102	97	73	33	41	25	3	14	1	1	4	4	4	3	7	4	13	14	25	8	17	3	1
5	강력범죄	강간	1155	334	201	352	203	176	97	17	77	4	22	19	24	27	18	36	8	87	93	128	55	108	29	6
6	강력범죄	유사강간	144	43	26	44	23	33	2	0	5	0	2	5	2	5	3	5	1	4	10	14	12	19	3	
7	강력범죄	강제추행	5582	1282	610	1282	490	481	302	53	258	24	83	68	98	75	99	161	30	342	298	454	150	288	43	23
8	강력범죄	기타강간	60	23	8	13	10	9	9	2	1	0	2	4	0	0	1	1	1	3	11	7	3	6	4	
9	강력범죄	방화	225	90	71	81	35	37	33	3	27	0	5	10	4	6	2	19	3	26	16	30	9	24	5	1
10	강력범죄	경도	41144	15807	8310	9216	4821	6216	3578	582	2488	202	1053	849	800	763	916	1359	350	4075	3131	5494	1293	2904	618	180
11	폭력범죄	상해	6016	4220	2781	2074	1171	1165	1409	152	540	27	156	175	151	114	185	322	96	540	622	842	263	453	140	40
12	폭력범죄	폭행	36312	10174	6885	11586	4856	5256	3863	285	2917	150	1165	1160	898	778	627	1698	375	2808	2959	4474	1596	2946	656	217
13	폭력범죄	채포감금	258	68	65	210	38	43	22	6	25	1	5	11	4	5	6	18	3	16	27	27	13	29	5	
14	폭력범죄	협박	2923	1010	688	928	362	503	420	52	303	19	81	104	61	51	80	183	39	276	348	528	213	370	47	18
15	폭력범죄	약취유인	49	18	8	15	3	6	8	0	4	0	1	3	1	0	1	3	0	4	2	10	2	4	0	
16	폭력범죄	폭력행위	2119	940	441	773	373	278	264	35	188	9	59	59	55	48	32	110	31	230	207	335	131	228	21	13
17	폭력범죄	강간	739	360	177	368	90	96	76	16	90	1	13	23	9	9	28	44	9	130	76	142	26	61	11	8
18	폭력범죄	손괴	10985	3707	2645	3456	1474	1463	1297	170	833	44	301	299	258	278	346	619	139	1093	924	1560	632	1182	254	60
19	지능범죄	직무남용	116	32	10	12	24	10	15	12	4	2	0	2	4	0	2	3	1	2	8	11	6	4	0	
20	지능범죄	직권남용	137	32	14	25	11	2	11	6	3	0	1	4	1	0	0	1	0	3	9	17	7	6	1	
21	지능범죄	중수뢰	54	25	10	13	6	5	4	0	4	1	1	1	0	0	1	3	1	10	0	8	1	3	2	
22	지능범죄	홍차	1146	76	40	152	27	37	24	9	65	11	11	16	30	3	8	53	6	39	98	64	7	26	15	4
23	지능범죄	온라인장	2590	829	446	579	336	346	225	50	170	11	58	58	52	41	65	104	18	180	211	225	65	191	35	11
24	지능범죄	유가증권	98	5	10	18	5	4	2	0	10	0	1	1	4	0	1	2	2	6	3	5	0	3	1	
25	지능범죄	사기	49935	19036	10495	12302	6450	6616	5198	785	3954	189	1091	1215	868	987	1714	1795	519	4927	3662	6250	2069	3636	742	220
26	지능범죄	횡령	15282	3376	2053	2879	1477	1287	863	252	1285	81	404	278	252	218	533	427	85	971	963	1324	319	844	164	78
27	지능범죄	배임	928	315	128	167	65	67	84	17	66	15	7	27	12	15	28	34	9	47	62	69	33	65	20	3
28	통속범죄	상속권	4102	1048	501	1060	360	451	216	60	219	26	59	57	55	48	70	85	23	324	310	445	123	286	18	20
29	통속범죄	도박	1216	1992	508	302	210	203	214	15	87	5	36	37	18	17	31	51	12	145	122	150	34	91	12	4
30	특별경제범죄	특별경제범죄	15101	4459	2103	3118	1265	1207	1042	230	933	26	555	276	256	312	269	511	145	911	1169	1133	507	805	143	59
31	마약범죄	마약범죄	1746	971	258	586	31	97	73	19	105	3	61	29	22	19	37	40	15	67	73	158	54	94	31	7
32	보안범죄	보안범죄	3606	1519	321	644	279	188	152	10	214	13	58	58	56	47	51	217	14	176	155	318	83	146	40	11
33	화재범죄	화재범죄	52	1196	14	79	7	12	34	1	63	4	9	75	11	4	24	97	7	24	17	32	31	27	3	

[사진15] 2017년 지역별 범죄발생 통계데이터



04. 연구 과정 및 결과

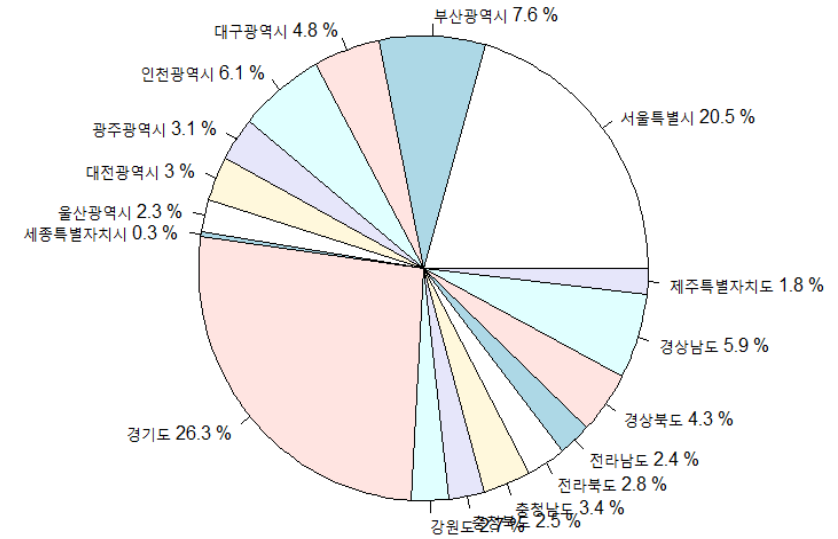
- 2017년 전국 시도별 범죄 관련 데이터 분석 후 시각화



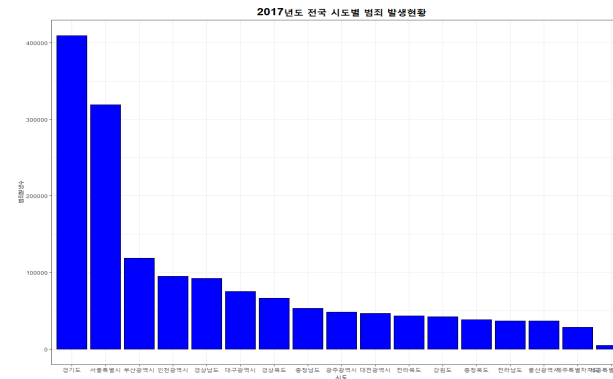
[사진18] 2017년도 전국 시도별 범죄 발생현황 지도시각화

분석 과정

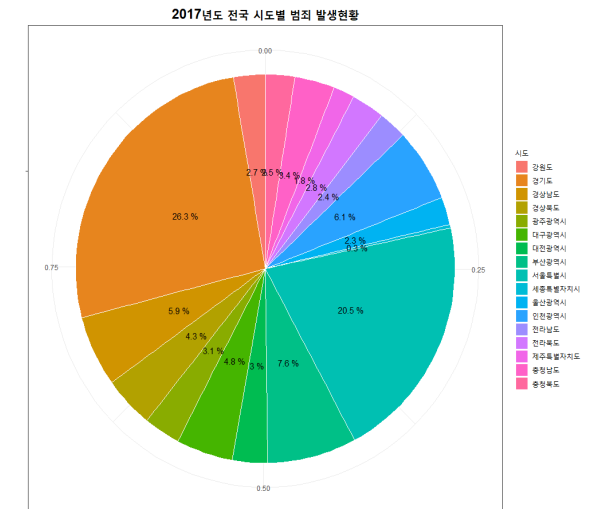
- 수집된 여러 데이터를 읽어옴
- 시각화에 요구되는 데이터로 가공
- ggplot2 패키지의 다양한 그래프로 시각화



[사진19] Pie Chart 시각화



[사진20] ggplot2 막대그래프 시각화

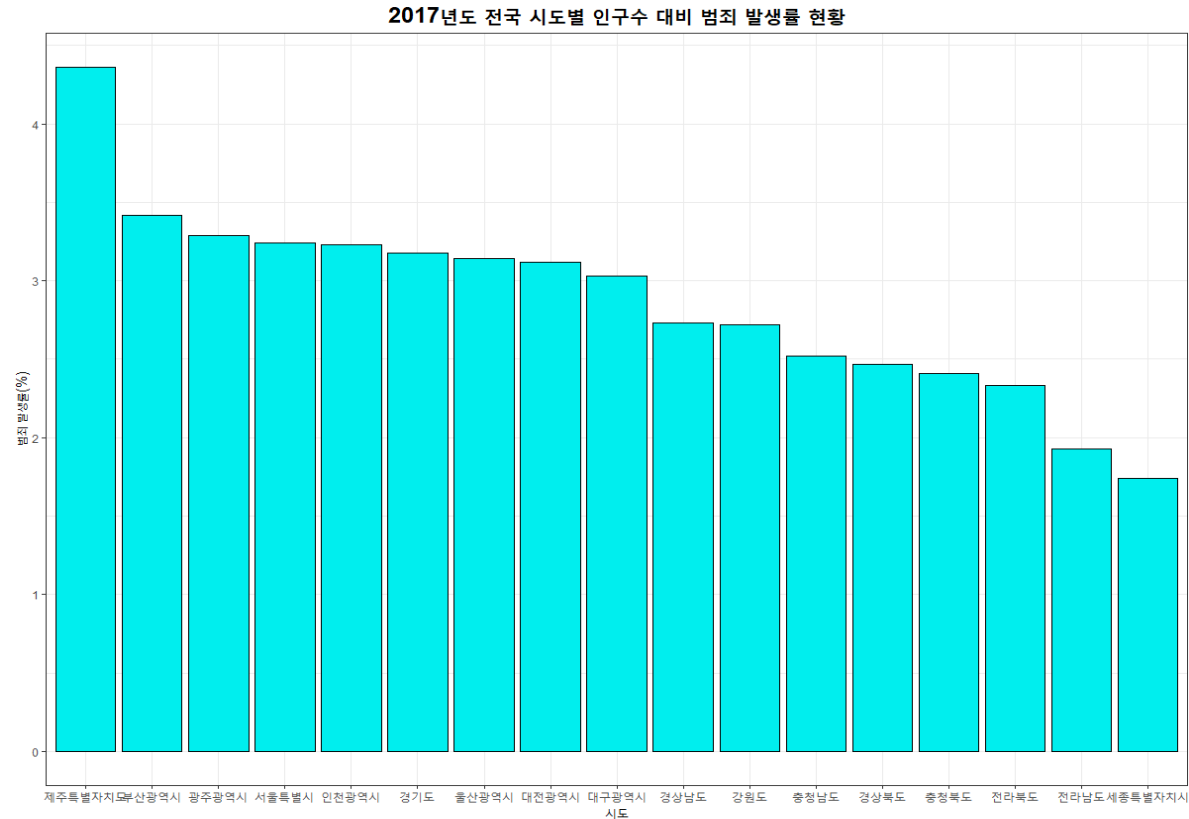


[사진21] ggplot2 원형그래프 시각화



04. 연구 과정 및 결과

- 2017년 전국 시도별 인구 수 대비 범죄 발생률 (%)



[사진22] 전국 시도별 인구수 대비 범죄 발생률 ggplot2 막대그래프 시각화

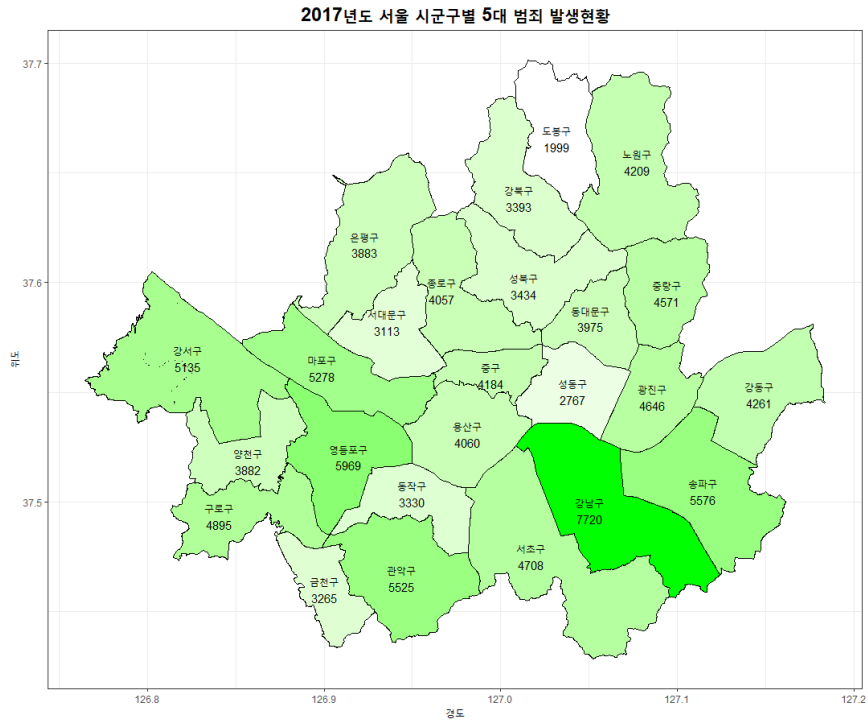
분석 과정

- 지역별 총 인구수 데이터 + 지역별 범죄 통계 데이터
- 시각화에 요구되는 데이터로 가공
- ggplot2 막대 그래프 시각화

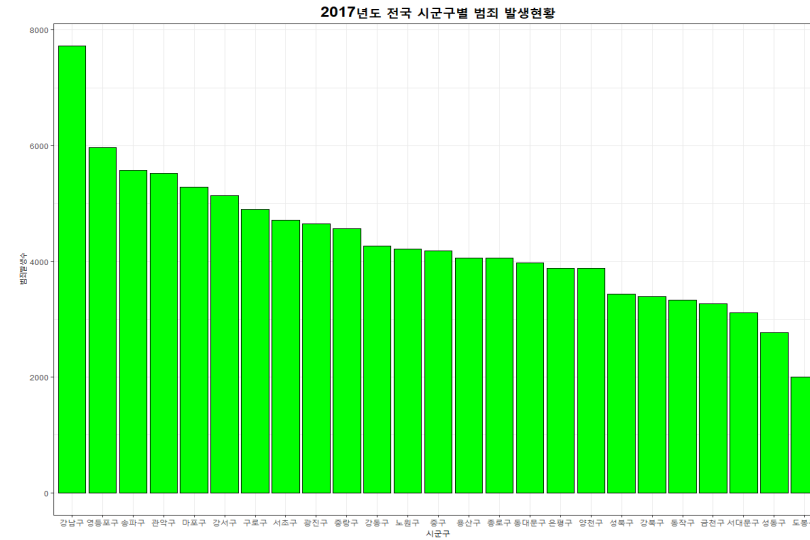


04. 연구 과정 및 결과

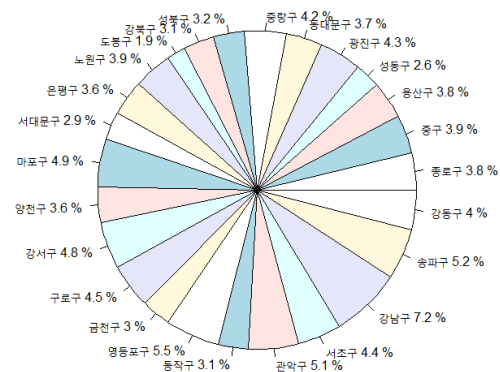
- 2017년 서울 시·군·구 별 5대 범죄 관련 데이터 분석 후 시각화



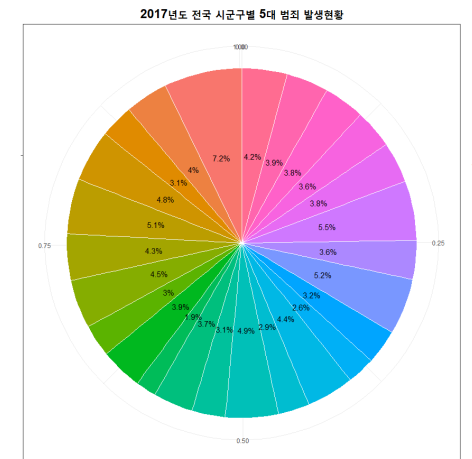
[사진23] 2017년도 서울 시군구별 범죄 발생현황 지도 시각화



[사진24] ggplot2 막대 그래프 시각화



[사진25] Pie Chart 시각화

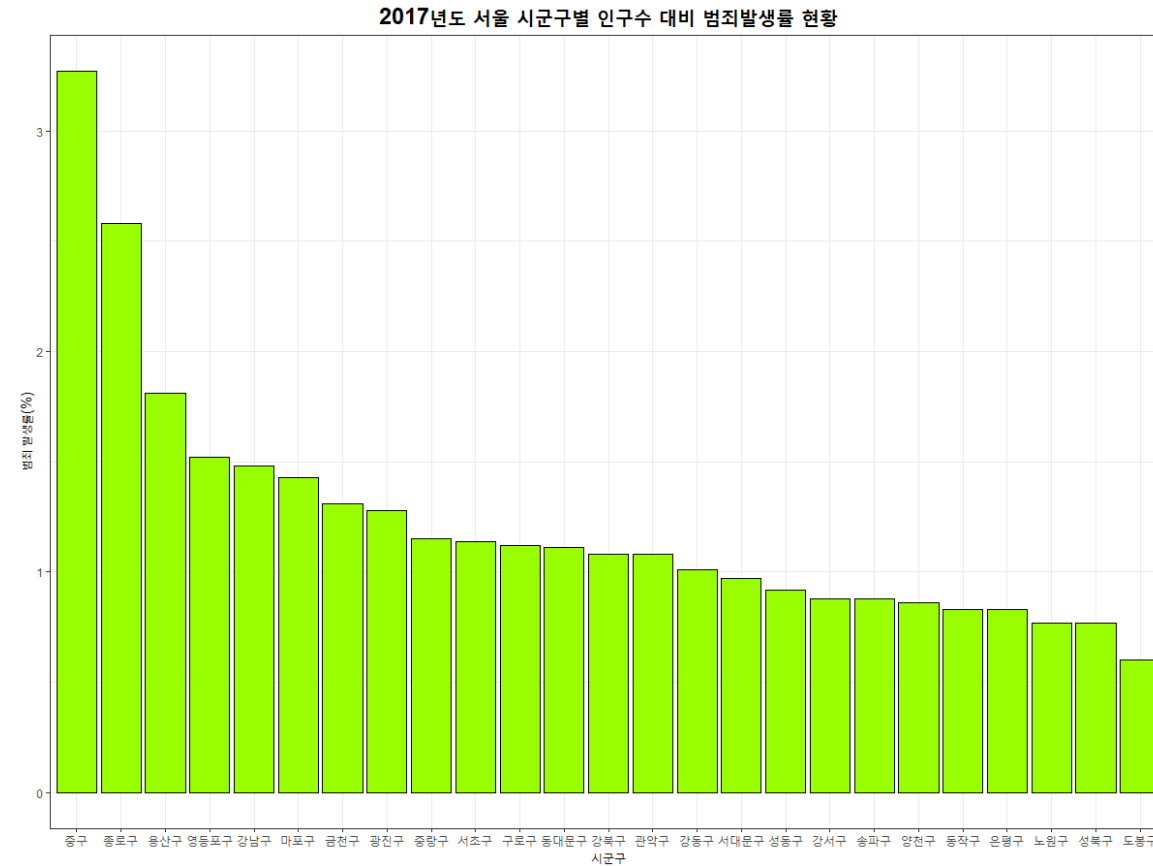


[사진26] ggplot2 원형 그래프 시각화



04. 연구 과정 및 결과

- 2017년 서울 시·군·구 별 인구수 대비 범죄 발생률 데이터 분석 후 시각화

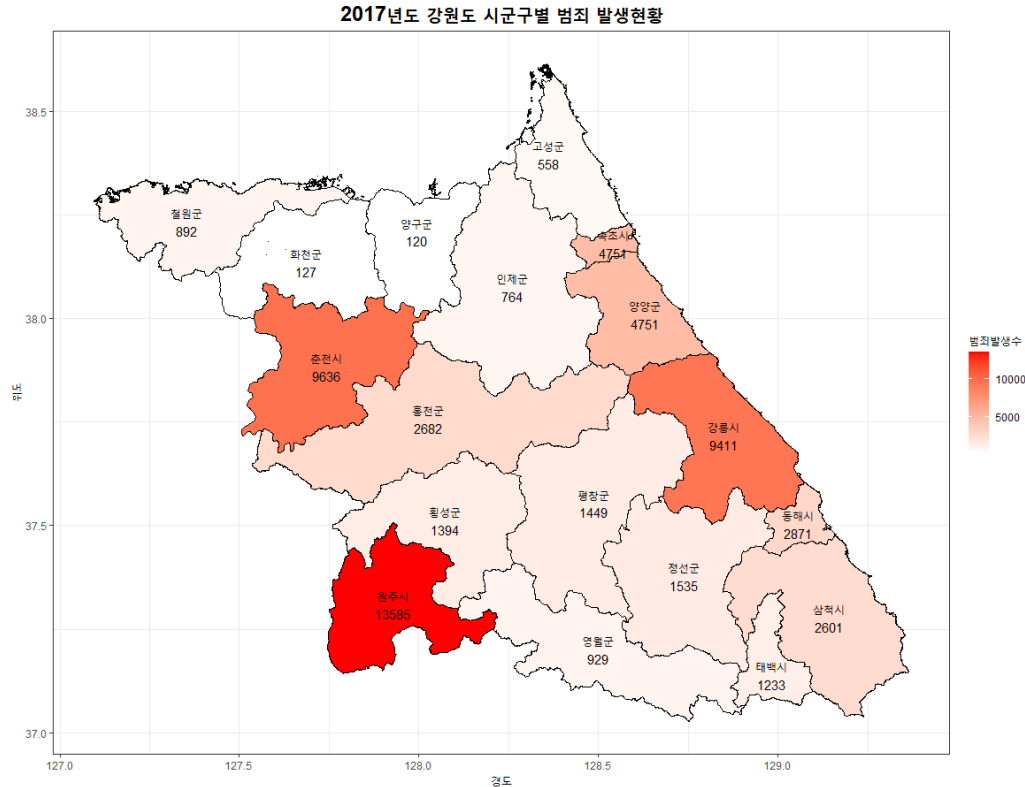


[사진27] 서울 시군구별 인구수 대비 범죄발생률 ggplot2 막대그래프 시각화

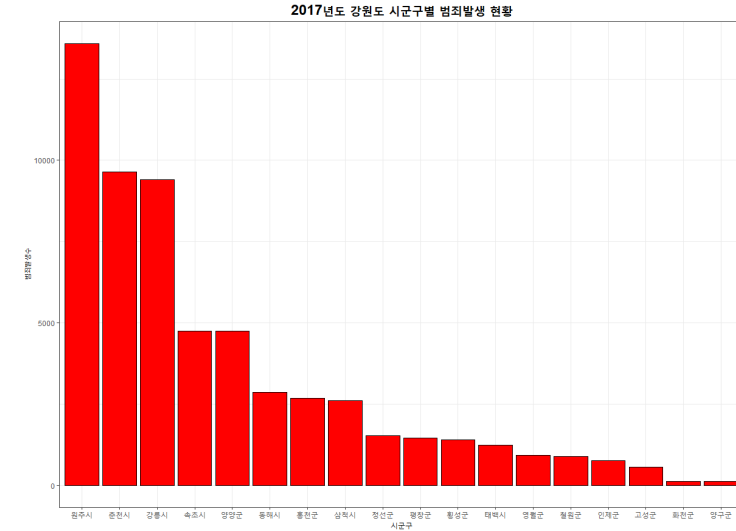


04. 연구 과정 및 결과

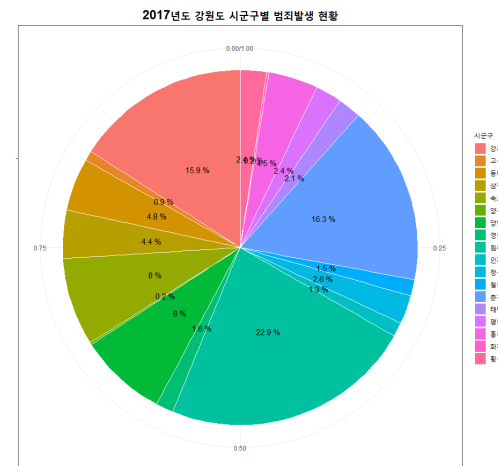
- 2017년 강원도 시·군·구 별 범죄 관련 데이터 분석 후 시각화



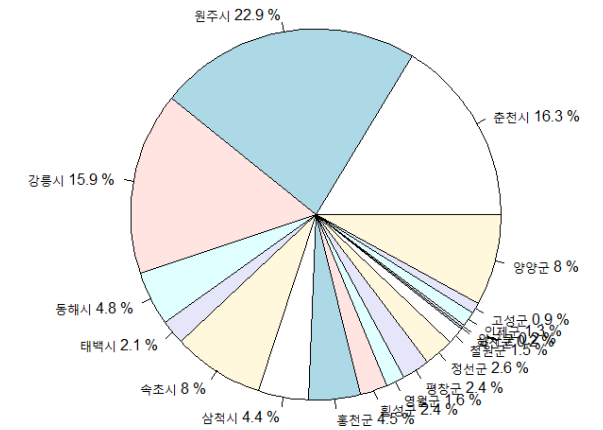
[사진28] 2017년도 강원도 시군구별 범죄 발생현황 지도시각화



[사진30] ggplot2 막대그래프 시각화



[사진31] 'ggplot2' 원형그래프 시각화

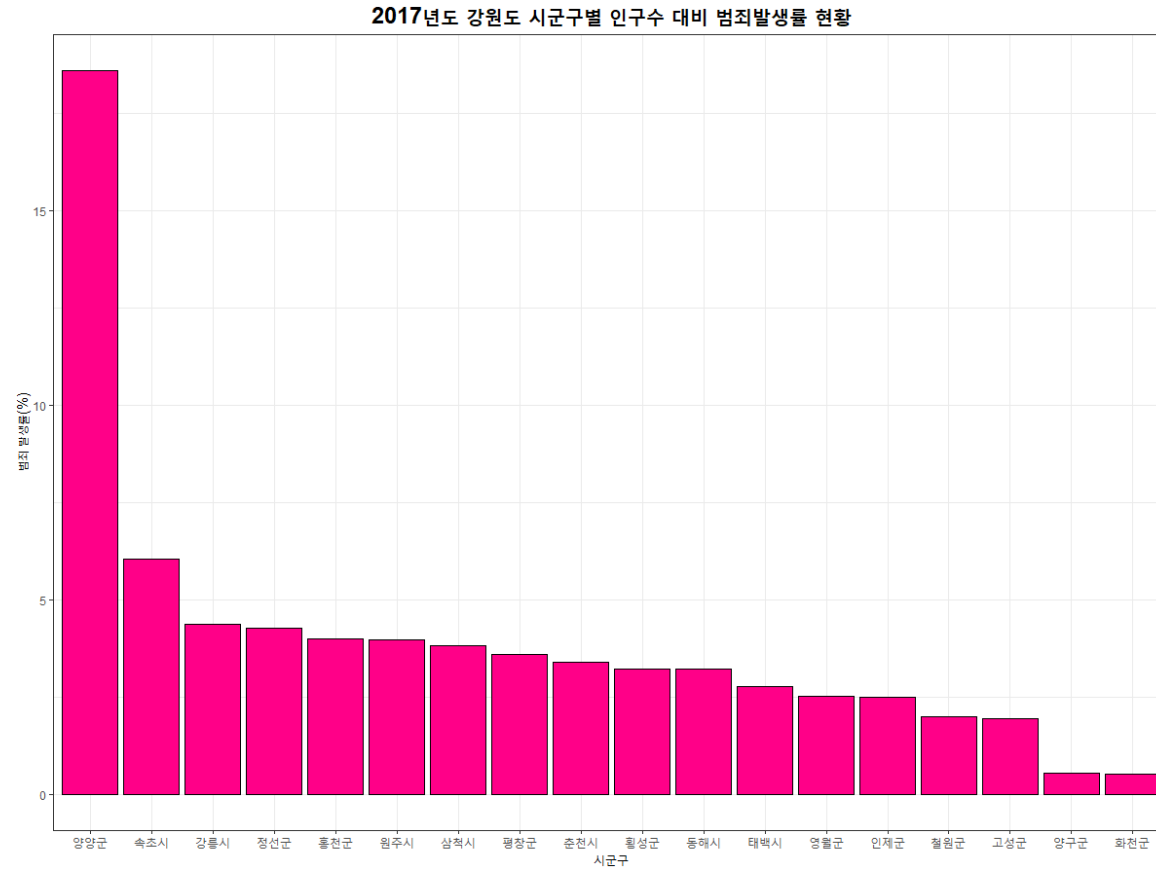


[사진31] Pie Chart 시각화



04. 연구 과정 및 결과

- 2017년 강원도 인구수 대비 범죄발생률 데이터 분석 후 시각화



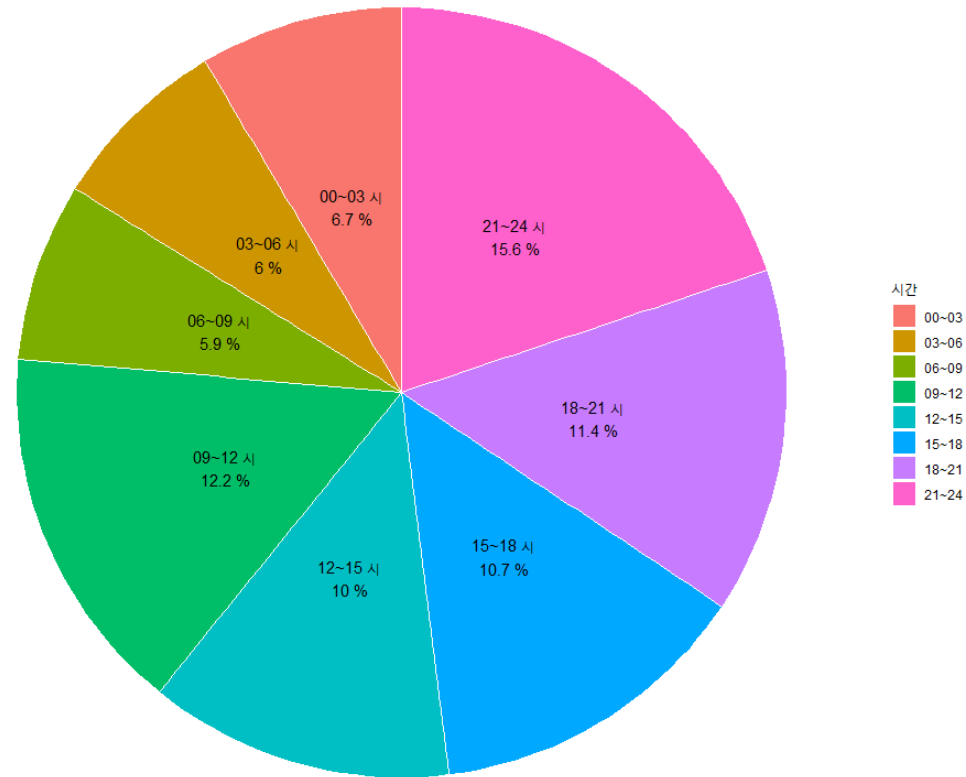
[사진32] 강원도 시군구별 인구수 대비 범죄발생률 ggplot2 막대그래프 시각화



04. 연구 과정 및 결과

- 2017년 전국 범죄 발생 시간 데이터 분석 후 시각화

2017년도 범죄 발생시간 현황



[사진33] 2017년도 범죄 발생시간 ggplot2 원 그래프 시각화

05. 시사점 및 한계점



05. 시사점 및 한계점

본 연구의 의의

모든 서비스 개발에는 **데이터 분석이 필수적**이므로, 이를 위해 데이터 분석 및 시각화를 학습하였다. 이후 연구하고자 하는 주제에 맞게 복잡한 과정을 거쳐 **데이터를 가공**하였으며, 가공한 데이터를 기반으로 다양하게 **시각화**를 함으로써 추후 본인의 개발과 기획에 활용할 수 있게 하였다. 또, 추후에 이를 서비스화 함으로써 타 연구에도 충분히 활용할 수 있는 가능성을 개발하였다.

다양한 분야로의 활용

본 연구는 여러 주제에 맞는 범죄 데이터를 이용하여 시각화함으로써 **다양한 연구 개발에 사용**될 수 있다. R 언어의 다양한 package를 활용하여, **범죄예방을 위한 어플리케이션**이나 **스마트 안전 도시 리빙랩** 등 다양한 분야에서 본 연구 자료를 활용할 수 있을 것이다. 더 나아가 원하는 데이터를 입력하면 자동으로 시각화 해주는 프로그램 등으로 발전시킬 수 있을 것으로 보인다.

본 연구의 한계점

통계청 자료와 다양한 데이터 수집 기관에서 데이터를 받아왔으나, 해당 데이터가 매우 **한정적**이었기 때문에 **원하는 데이터를 정확히 받아오기 힘들었다**는 한계가 있었다. 이를 극복하기 위해 **직접 다양한 데이터셋을 만들어 분석**하는 등의 노력을 하였다. 이후 기존에 학교에서 학습한 다양한 지식을 이용하여 플랫폼을 제작하는 등의 추가 연구도 진행하고 싶다.

06. Q&A

감 사 합 니 다