L14. Geospatial data (1)

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

SEOULTECH 서울과학기술대학교 데이터사이언스학과

I. Introduction

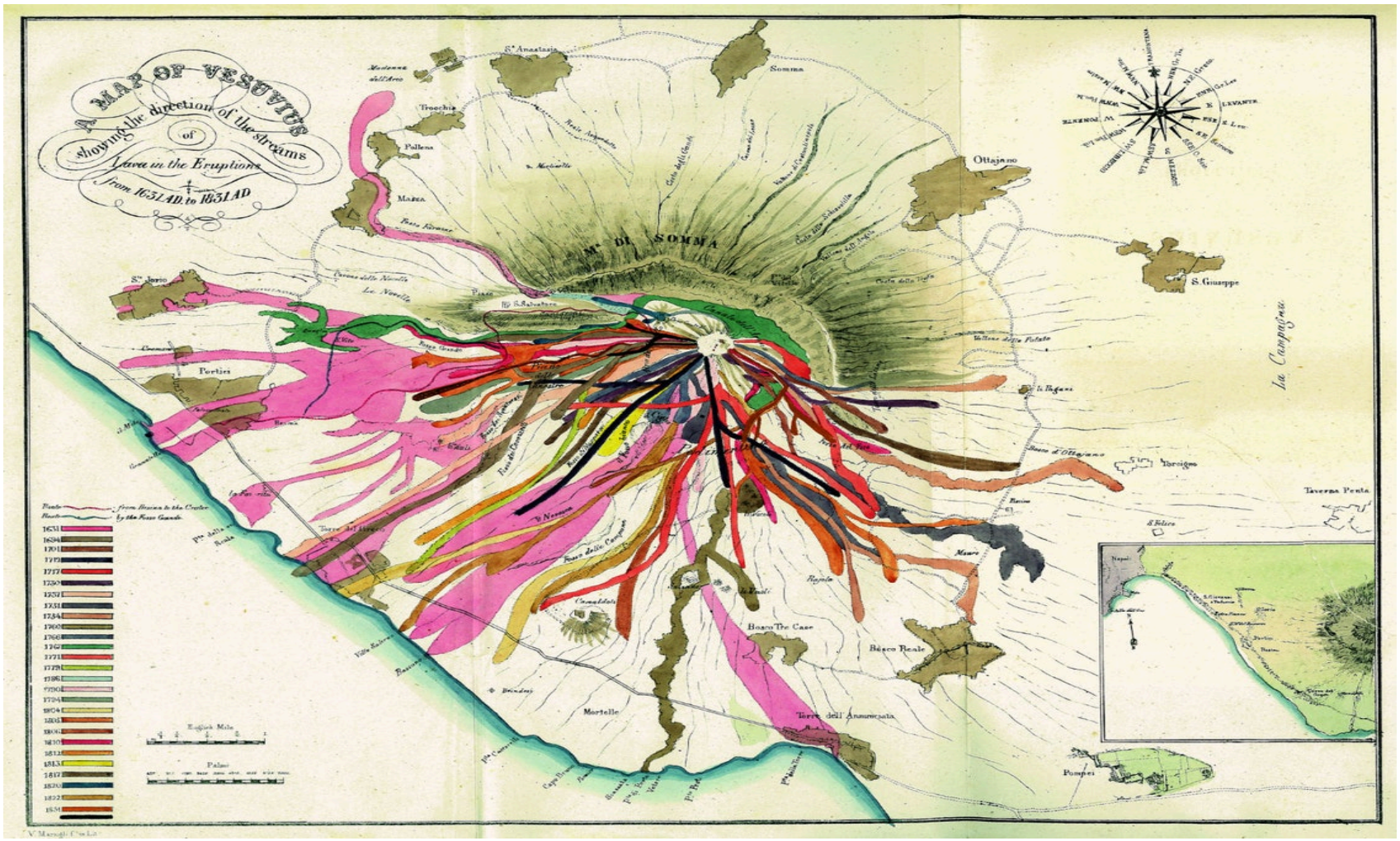
지리정보의 중요성

- 지도는 기원전 6세기부터 제작되어 인간의 삶에 뗄수 없는 중요한 도구
- 지리 정보와 연관된 질문들
  - 어디에 점포를 개점할 것인가?
  - 어디에 물류 센터를 위치시킬 것인가?
  - 어느지역이 공공시설이 더 필요한가?
  - 어디로 이사갈 것인가?
  - ...
- GPS, Satellite, CCTV등의 관련 인프라가 확산됨에 따라서 Geographical Information System (GIS)는 최근 10년간 급격히 발전하여 하나의 독립적인 필드로 자리잡음

Cartography (지도학)

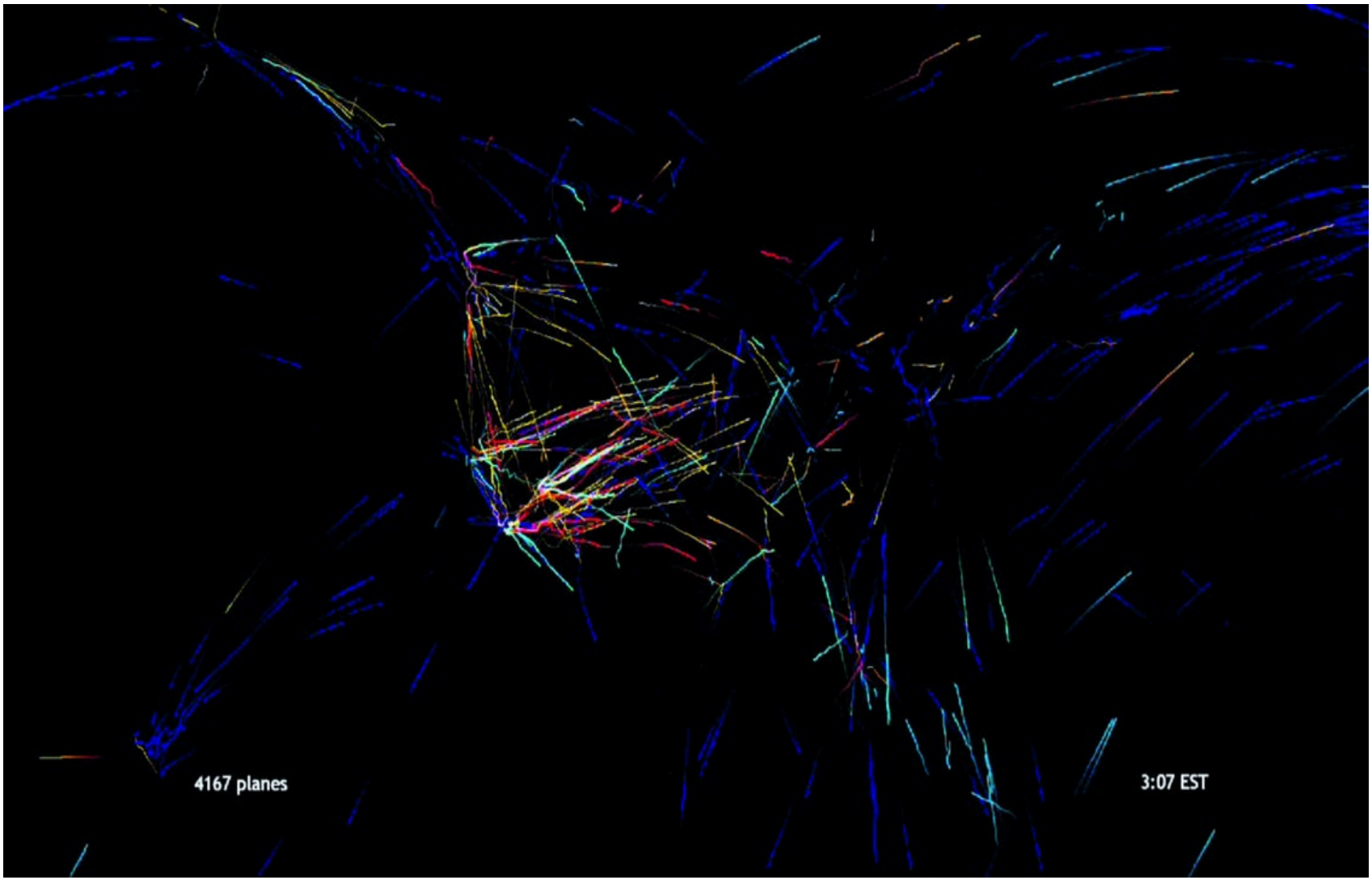
- 지도를 그리는 것에 대한 학문
- 지도에 대한 기술은 문명의 발전과 함께 발전함
- https://en.wikipedia.org/wiki/Cartography
- [8 Stunning Maps That Changed Cartography]
  - https://www.wired.com/2015/10/8-stunning-maps-changed-cartography/

John Auldjo's Map of Vesuvius was drawn in 1832 and depicts the path of lava flows.



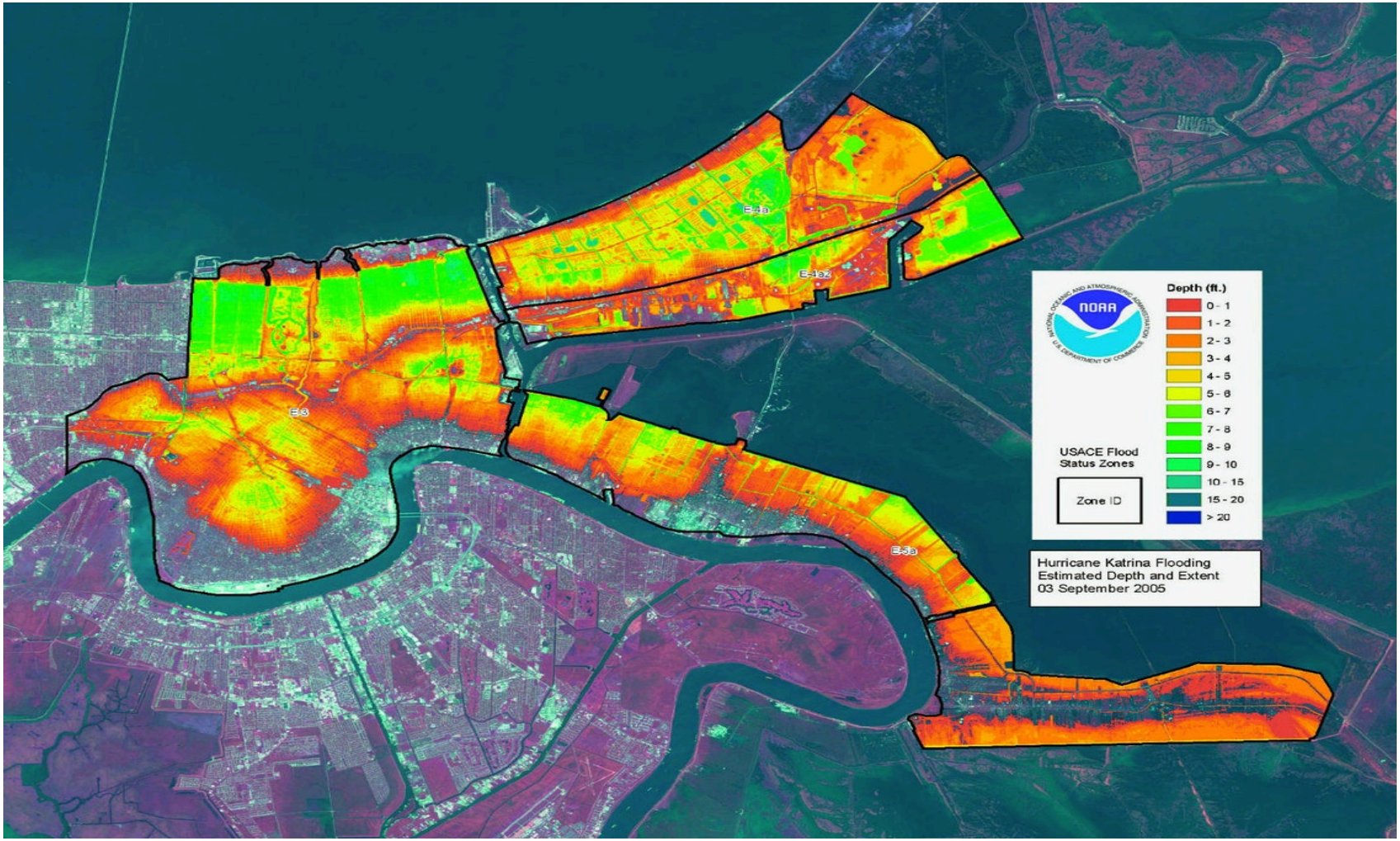
<Source: 8 Stunning Maps That Changed Cartography>

Aaron Koblin's Flight Paths map



<Source: 8 Stunning Maps That Changed Cartography>

This illustration of the flooding from Hurricane Katrina was created using the best technology of the day.

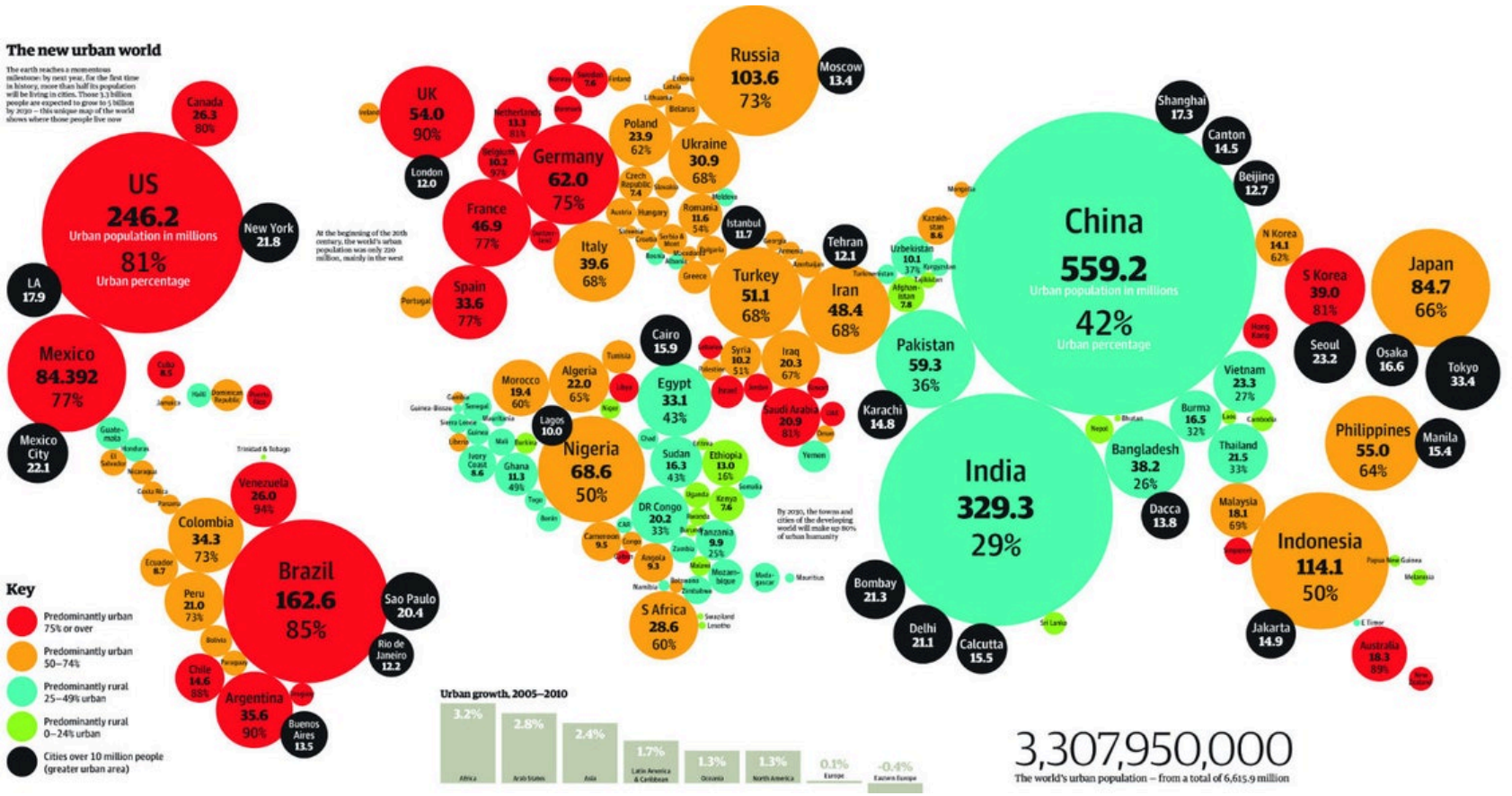


<Source: 8 Stunning Maps That Changed Cartography>

This world map from 1482 was based on the writings from Claudius Ptolemy, an ancient geographer who is credited with defining latitude and longitude.Daniel Crouch Rare Books

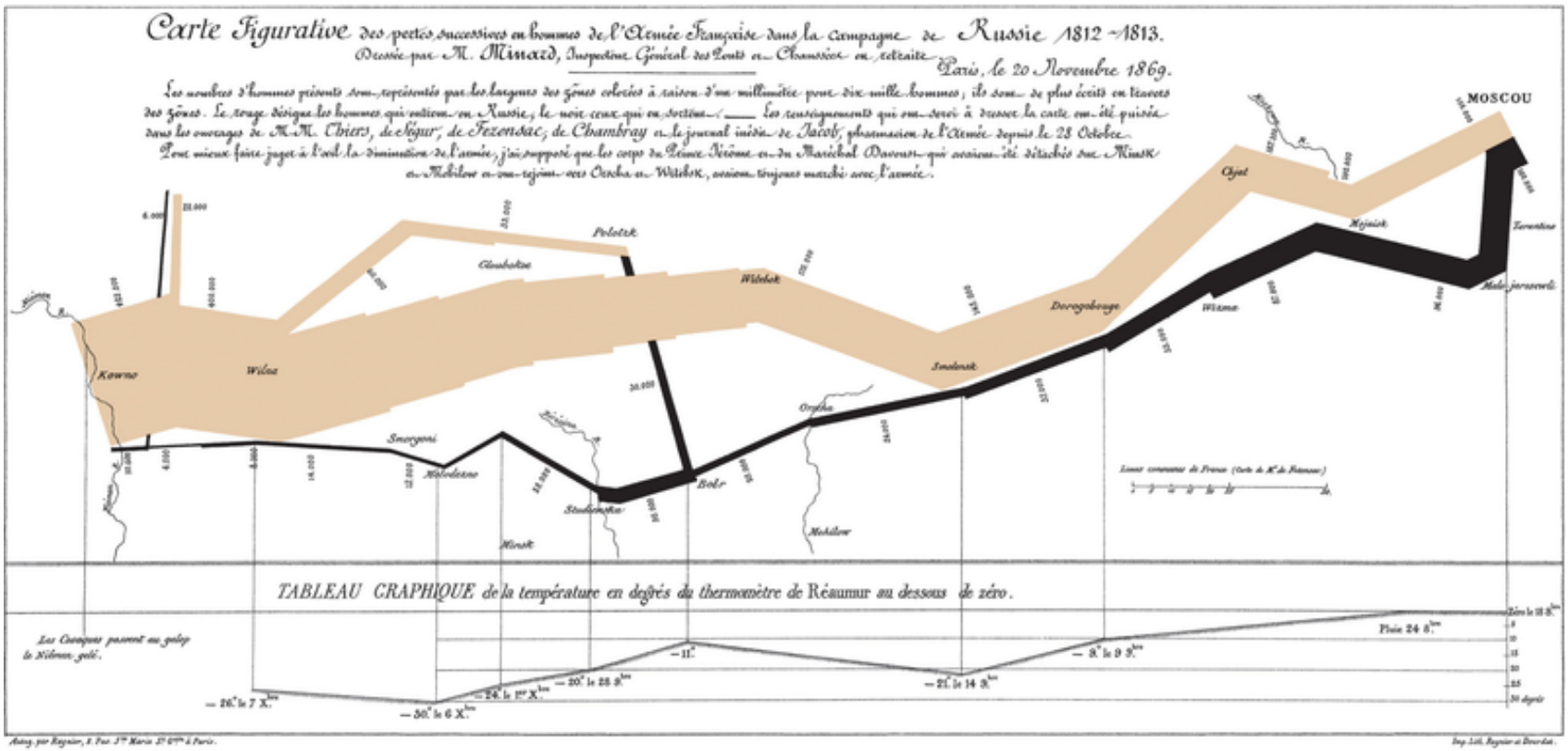
<Source: 8 Stunning Maps That Changed Cartography>

This map from Paul Scruton depicts the population of countries as circles. It's a good example of how the lines between info graphics and cartography are blurring.



<Source: 8 Stunning Maps That Changed Cartography>

Charles Minard's map of Napoleon's disastrous Russian campaign of 1812. The graphic is notable for its representation in two dimensions of six types of data: 1) the number of Napoleon's troops; 2) distance; 3) temperature; 4) the latitude and longitude; 5) direction of travel; and 5) location relative to specific dates.

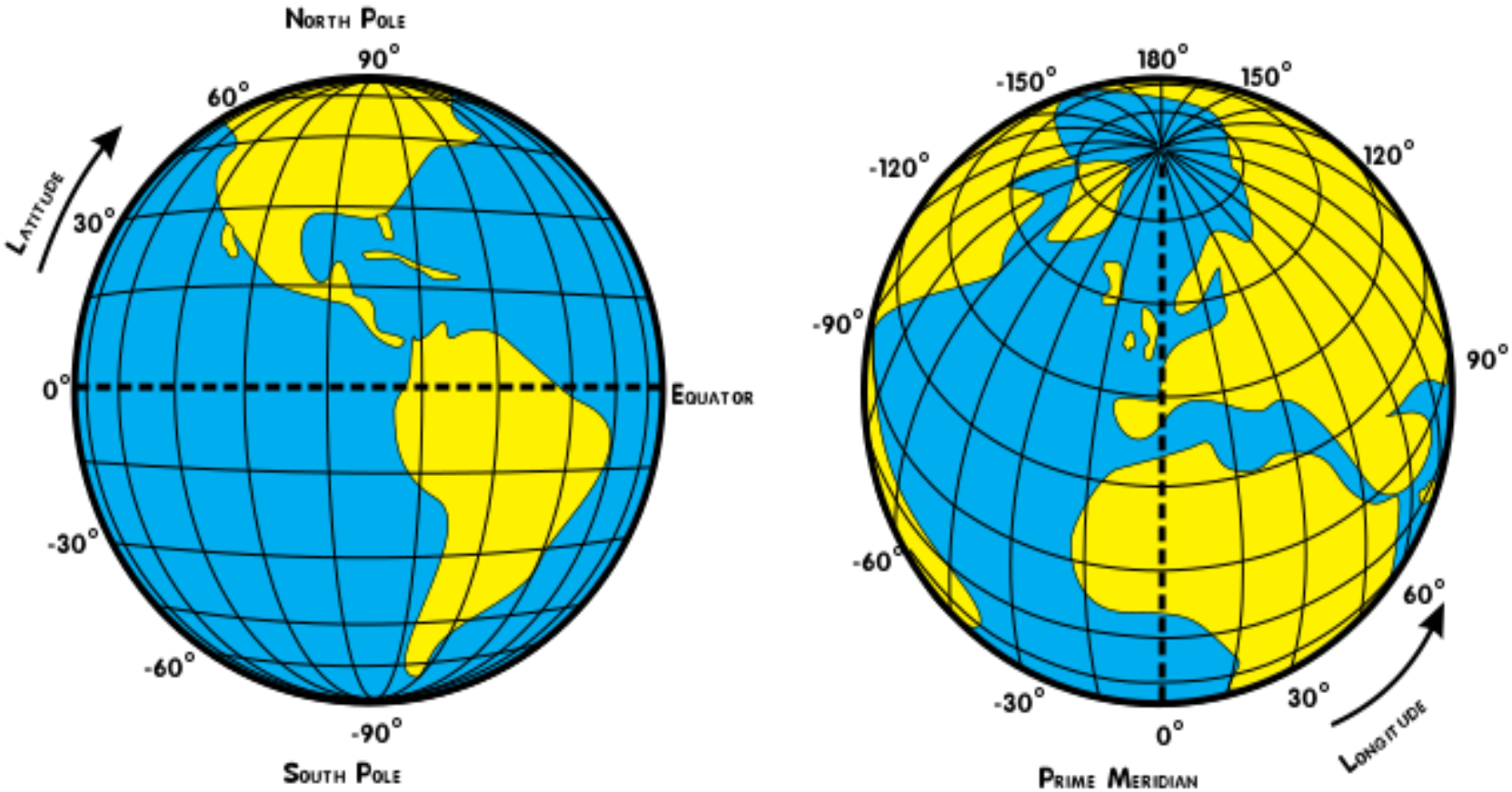


<Source: Wikipedia - Charles_Joseph_Minard>

Takeaways

1. Latitude (위도, 남북) and Longitude (경도, 동서) are two essential aesthetic elements.
2. Dynamic data can be described with careful implementation.
3. Modern cartography involves modern technology such as satellite image.
4. The border between graphics and cartography is blurring.

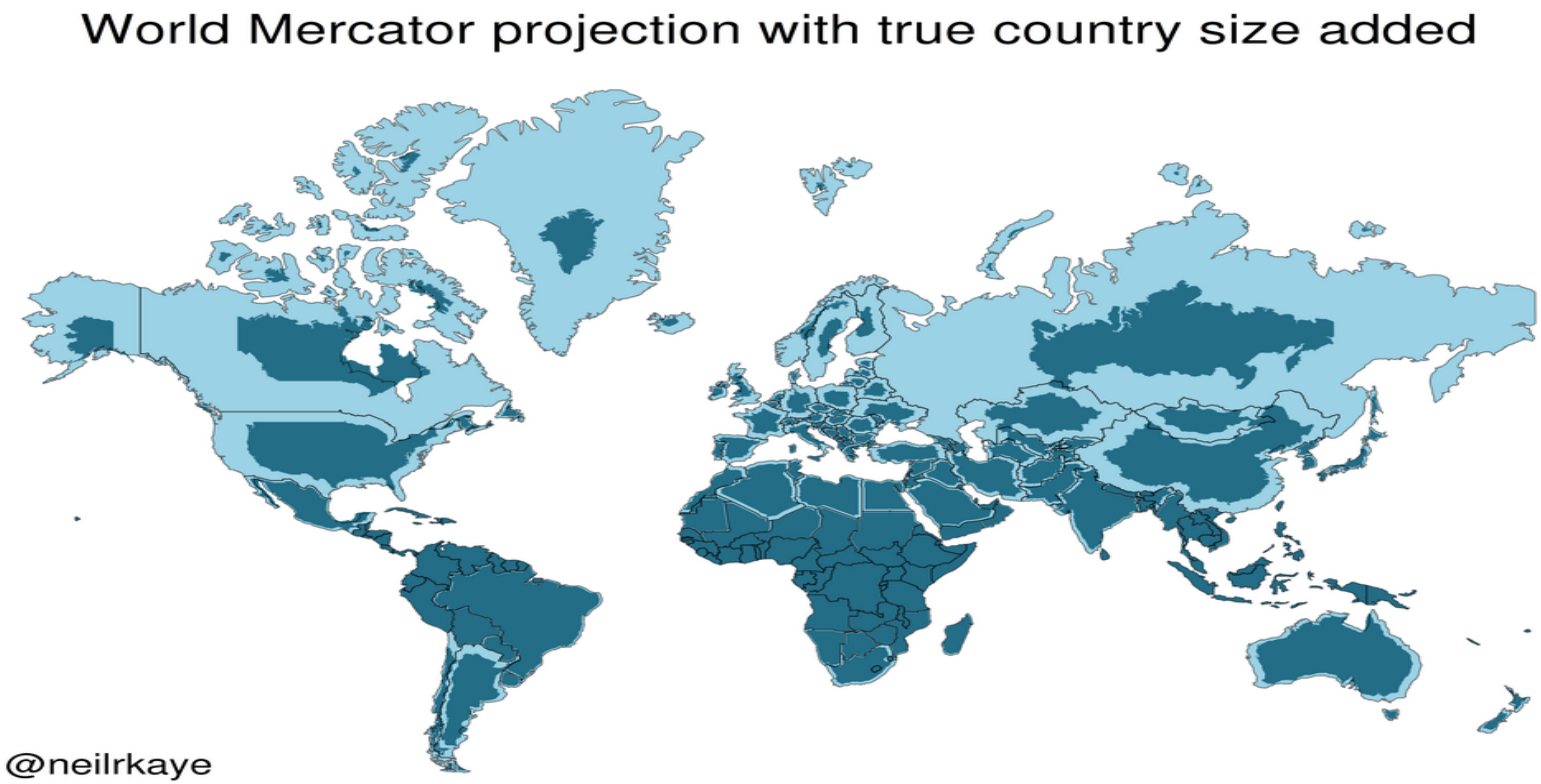
Latitude (위도) & Longitude (경도)



<Source: Wikipedia Commons - Latitude_and_Longitude_of_the_Earth.svg>

2D 지도의 왜곡

지도에서는 구형의 지구를 2D로 표현하기 때문에 왜곡이 불가피함



<Source: Reddit.com - /r/dataisbeautiful/comments/9nkg7k/map_projections_can_be_deceptive_oc/>

II. Dataset for this module

kr_region_pop.xlsx

- 2015-2018년의 지역별 남성, 여성, 외국인 인구 (전국 sheet)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		2015	2015	2015	2015	2016	2016	2016	2016	2017	2017	2017	2017	2018	2018	2018	2018
2		총인구 (명)	남자 (명)	여자 (명)	외국인-계 (명)	총인구 (명)	남자 (명)	여자 (명)	외국인-계 (명)	총인구 (명)	남자 (명)	여자 (명)	외국인-계 (명)	총인구 (명)	남자 (명)	여자 (명)	외국인-계 (명)
3	전국	51,069,375	25,608,502	25,460,873	1,363,712	51,269,554	25,696,987	25,572,567	1,413,758	51,422,507	25,768,055	25,654,452	1,479,247	51,629,512	25,877,195	25,752,317	1,651,561
4	서울특별시	9,904,312	4,859,535	5,044,777	337,116	9,805,506	4,799,115	5,006,391	335,167	9,741,871	4,757,642	4,984,229	343,927	9,673,936	4,718,972	4,954,964	374,425
5	부산광역시	3,448,737	1,701,347	1,747,390	44,070	3,440,484	1,694,026	1,746,458	46,168	3,416,918	1,680,933	1,735,985	48,808	3,395,278	1,668,984	1,726,294	55,506
6	대구광역시	2,466,052	1,228,511	1,237,541	29,282	2,461,002	1,223,733	1,237,269	30,492	2,453,041	1,218,326	1,234,715	31,365	2,444,412	1,212,777	1,231,635	35,280
7	인천광역시	2,890,451	1,455,017	1,435,434	67,850	2,913,024	1,465,699	1,447,325	71,873	2,925,967	1,472,081	1,453,886	79,170	2,936,117	1,476,097	1,460,020	88,489
8	광주광역시	1,502,881	748,867	754,014	21,592	1,501,557	747,303	754,254	23,701	1,496,172	745,122	751,050	26,052	1,490,092	741,728	748,364	29,833
9	대전광역시	1,538,394	772,243	766,151	19,080	1,535,445	770,971	764,474	19,687	1,525,849	765,718	760,131	20,534	1,511,214	758,035	753,179	22,858
10	울산광역시	1,166,615	606,924	559,691	29,860	1,166,033	605,618	560,415	28,792	1,157,077	600,093	556,984	26,601	1,150,116	595,795	554,321	27,162
11	세종특별자치시	204,088	103,210	100,878	4,471	242,507	122,648	119,859	4,835	276,589	139,347	137,242	5,290	312,374	157,289	155,085	6,486
12	경기도	12,479,061	6,309,661	6,169,400	452,632	12,671,956	6,405,301	6,266,655	472,699	12,851,601	6,502,723	6,348,878	497,089	13,103,188	6,635,106	6,468,082	558,197
13	강원도	1,518,040	768,241	749,799	18,306	1,521,751	769,461	752,290	19,364	1,521,386	769,043	752,343	20,996	1,520,391	768,604	751,787	24,149
14	충청북도	1,589,347	805,377	783,970	40,758	1,603,404	814,049	789,355	44,901	1,611,009	818,500	792,509	48,237	1,620,935	826,183	794,752	54,835
15	충청남도	2,107,802	1,078,310	1,029,492	71,082	2,132,566	1,091,091	1,041,475	77,533	2,162,426	1,108,610	1,053,816	84,752	2,181,416	1,119,996	1,061,420	95,838
16	전라북도	1,834,114	915,729	918,385	29,930	1,833,168	915,493	917,675	31,549	1,826,174	911,592	914,582	33,764	1,818,157	907,683	910,474	39,131
17	전라남도	1,799,044	900,967	898,077	34,611	1,796,017	901,500	894,517	36,630	1,792,319	899,077	893,242	37,212	1,790,352	900,154	890,198	42,520
18	경상북도	2,680,294	1,351,037	1,329,257	57,565	2,682,169	1,354,997	1,327,172	60,706	2,677,058	1,353,083	1,323,975	63,978	2,672,902	1,352,869	1,320,033	74,071
19	경상남도	3,334,524	1,698,737	1,635,787	90,361	3,339,633	1,701,849	1,637,784	92,185	3,345,293	1,702,150	1,643,143	91,072	3,350,350	1,703,904	1,646,446	97,494
20	제주특별자치도	605,619	304,789	300,830	15,146	623,332	314,133	309,199	17,476	641,757	324,015	317,742	20,400	658,282	333,019	325,263	25,287
21																	
22																	
23																	

전국 서울

- 2015-2018년의 서울시 구별 남성, 여성, 외국인 인구 (서울 sheet)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		2015	2015	2015	2015	2016	2016	2016	2016	2017	2017	2017	2017	2018	2018	2018	2018
2		총인구 (명	남자 (명)	여자 (명)	외국인-계	총인구 (명	남자 (명)	여자 (명)	외국인-계	총인구 (명	남자 (명)	여자 (명)	외국인-계	총인구 (명	남자 (명)	여자 (명)	외국인-계 (명)
3	종로구	161,521	79,510	82,011	15,402	154,398	75,620	78,778	10,349	157,277	76,670	80,607	10,979	157,967	77,147	80,820	12,056
4	중구	128,478	63,218	65,260	10,457	128,160	62,514	65,646	10,200	127,896	62,195	65,701	10,265	129,797	62,975	66,822	11,988
5	용산구	227,282	109,980	117,302	15,093	224,993	108,520	116,473	15,283	223,898	108,471	115,427	16,005	226,938	110,685	116,253	19,126
6	성동구	295,006	146,332	148,674	9,869	294,744	145,935	148,809	10,009	302,367	149,060	153,307	10,436	306,796	150,623	156,173	11,307
7	광진구	368,199	180,647	187,552	17,206	363,979	177,858	186,121	17,503	363,934	177,156	186,778	18,399	362,304	175,762	186,542	19,729
8	동대문구	364,787	181,189	183,598	14,830	359,935	178,285	181,650	16,191	357,380	176,460	180,920	17,942	358,141	176,323	181,818	19,848
9	중랑구	403,237	199,936	203,301	5,750	399,317	197,875	201,442	5,770	396,892	196,346	200,546	6,013	391,668	193,599	198,069	6,586
10	성북구	456,844	221,861	234,983	11,134	451,800	218,606	233,194	12,177	445,417	215,068	230,349	13,191	438,734	211,204	227,530	14,523
11	강북구	319,992	156,386	163,606	4,238	316,212	154,407	161,805	4,295	313,698	152,464	161,234	4,443	309,138	150,028	159,110	4,947
12	도봉구	340,095	166,187	173,908	2,554	336,745	164,407	172,338	2,580	332,586	161,944	170,642	2,629	328,243	159,569	168,674	2,951
13	노원구	562,996	274,128	288,868	4,726	555,420	270,058	285,362	4,691	543,499	263,655	279,844	4,731	534,096	258,737	275,359	5,828
14	은평구	478,374	232,709	245,665	5,599	472,242	229,215	243,027	5,429	466,243	225,813	240,430	5,592	462,552	223,452	239,100	5,970
15	서대문구	308,768	147,956	160,812	11,007	320,258	151,756	168,502	11,839	321,345	151,132	170,213	13,152	318,874	149,438	169,436	14,366
16	마포구	381,330	182,657	198,673	11,455	375,142	179,193	195,949	11,401	368,841	175,326	193,515	11,960	368,181	174,194	193,987	12,776
17	양천구	465,512	229,878	235,634	5,847	460,267	226,927	233,340	5,646	452,111	222,503	229,608	5,468	445,591	218,975	226,616	5,681
18	강서구	570,507	278,270	292,237	8,039	574,287	279,521	294,766	8,319	581,675	282,678	298,997	8,682	578,539	280,629	297,910	9,179
19	구로구	444,832	222,985	221,847	39,461	442,165	221,227	220,938	40,530	436,869	218,428	218,441	41,907	433,765	217,086	216,679	45,686
20	금천구	250,690	128,036	122,654	24,792	247,819	126,271	121,548	25,029	249,930	127,252	122,678	25,443	249,344	127,278	122,066	27,524
21	영등포구	406,528	203,801	202,727	49,044	398,120	199,142	198,978	47,757	393,560	196,929	196,631	46,708	395,286	198,031	197,255	49,151
22	동작구	407,894	200,602	207,292	14,925	404,076	197,854	206,222	15,412	400,236	195,253	204,983	16,009	397,980	193,713	204,267	17,321
23	관악구	519,622	261,967	257,655	24,022	515,648	259,824	255,824	24,466	511,222	256,857	254,365	24,475	510,303	255,687	254,616	25,418
24	서초구	420,804	201,841	218,963	7,109	419,682	201,062	218,620	6,667	414,550	198,241	216,309	6,538	409,491	196,064	213,427	7,143
25	강남구	541,688	259,526	282,162	8,627	532,469	254,645	277,824	7,985	522,514	250,087	272,427	7,655	507,810	243,579	264,231	8,814
26	송파구	634,941	309,336	325,605	9,746	628,212	305,494	322,718	9,746	633,953	307,723	326,230	9,418	638,167	309,442	328,725	10,123
27	강동구	444,385	220,597	223,788	6,184	429,416	212,899	216,517	5,893	423,978	209,931	214,047	5,887	414,231	204,752	209,479	6,384
28																	
29																	
30																	
31																	

전국 서울

전국 *sheet*

```
library(readxl)
kr_pop_header <-
  read_excel("data/kr_region_pop.xlsx",
             sheet = 1,
             col_names = FALSE)[1:2,]
kr_pop <-
  read_excel("data/kr_region_pop.xlsx",
             sheet = 1, skip = 2,
             col_names = FALSE)
colnames(kr_pop) <-
  c("state",
    paste(kr_pop_header[1,-1],
          kr_pop_header[2,-1], sep="_"))
kr_pop <- kr_pop %>%
  gather(colnames(kr_pop)[-1],
         key="var", value="pop") %>%
  separate(var,
           into=c("year", "category"),
           sep="_")
```

```
set.seed(123)
kr_pop %>% sample_n(10)
```

```
## # A tibble: 10 x 4
##    state       year  category       pop
##    <chr>       <chr> <chr>          <dbl>
##  1 강원도      2016  총인구 (명)   1521751
##  2 강원도      2018  총인구 (명)   1520391
##  3 세종특별자치시 2016  여자 (명)     119859
##  4 제주특별자치도 2018  남자 (명)     333019
##  5 경상북도    2018  여자 (명)     1320033
##  6 충청남도    2015  총인구 (명)   2107802
##  7 인천광역시  2017  총인구 (명)   2925967
##  8 경상남도    2018  남자 (명)     1703904
##  9 강원도      2017  총인구 (명)   1521386
##  10 서울특별시          2016    외국인-
계 (명)  335167
```

```
kr_pop %>%
  write.csv("data/kr_pop_tidy.csv",
            fileEncoding = 'euc-kr')
```

서울 *sheet*

```
library(readxl)
seoul_pop_header <-
  read_excel("data/kr_region_pop.xlsx",
             sheet = 2,
             col_names = FALSE)[1:2,]
seoul_pop <-
  read_excel("data/kr_region_pop.xlsx",
             sheet = 2, skip = 2,
             col_names = FALSE)
colnames(seoul_pop) <-
  c("district",
    paste(seoul_pop_header[1,-1],
          seoul_pop_header[2,-1], sep="_"))
seoul_pop <- seoul_pop %>%
  gather(colnames(seoul_pop)[-1],
         key="var",
         value="pop") %>%
  separate(var,
           into=c("year", "category"),
           sep="_")
```

```
set.seed(123)
seoul_pop %>% sample_n(10)
```

```
## # A tibble: 10 x 4
##   district    year  category       pop
##   <chr>       <chr> <chr>        <dbl>
## 1       강서구  2016  총인구 (명)   574287
## 2       양천구  2018  총인구 (명)   445591
## 3       서대문구 2016  여자 (명)    168502
## 4       종로구  2018  여자 (명)     80820
## 5       강남구  2018  여자 (명)    264231
## 6       금천구  2015  총인구 (명)   250690
## 7       강북구  2017  총인구 (명)   313698
##    8        서초구      2018    외국인-
계 (명)   7143
## 9       구로구  2017  총인구 (명)   436869
##  10        성동구      2016    외국인-
계 (명)  10009
```

```
seoul_pop %>%
  write.csv("data/seoul_pop_tidy.csv",
            fileEncoding = 'euc-kr')
```

kr_latlon.csv

```
kr_latlon <- read_csv(
  "data/kr_latlon.csv",
  locale = locale('ko', encoding='euc-kr'))
head(kr_latlon)
```

```
## # A tibble: 6 x 4
##   area             lat   lon note
##   <chr>          <dbl> <dbl> <chr>
## 1 서울특별시      37.6  127. <NA>
## 2 제주특별자치도  33.5  127. <NA>
## 3 전라남도        34.8  128. 여수
## 4 전라북도        35.8  127. 전주
## 5 광주광역시      35.2  127. <NA>
## 6 경상남도        35.2  128. 진주
```

```
tail(kr_latlon)
```

```
## # A tibble: 6 x 4
##   area       lat   lon note
##   <chr>    <dbl> <dbl> <chr>
## 1 서대문구  37.6  127. <NA>
## 2 양천구    37.5  127. <NA>
## 3 영등포구  37.5  127. <NA>
## 4 관악구    37.5  127. <NA>
## 5 성동구    37.6  127. <NA>
## 6 용산구    37.5  127. <NA>
```

```
"Hello"
```

```
## [1] "Hello"
```