Topics Dive: Legacy for the Future

- Shaokang Li

Introduction:

During past decades, numerous data storage devices appeared and dominated the market at some timestamp. For example, the VHS format, developed by JVC, is released in 1976, and in 2016, no one would make VHS products since there are much more advanced storage devices (DVD, Blu-Ray etc.). Instead of using physical media, people would stream contents on their laptop or smartphone with access to the Internet.

VHS Cassette

In this topic dive, I would like to look at questions below, exploring facts about these legacy items.

- What's the trend of different legacy items?
- Will anyone still checkout these legacy today?
    - If so, which item do they borrow?
        - Are there any findings in these item?

Description:

The first step is to find out what storage devices the Seattle Public Library has.

Go to website below, we may find out what's the itemType for these legacy devices.

These media/data storage devices are mainly categorized as Equipment and Media. Here's some of these devices with its itemType code.

- USB Floppy Drive (`ucflpdr` , `naflpdr` , `acdisk` , `jrdisk` , `ardisk`)
- VHS (`bcvhs` , `jcvhs` , `navid` , `alvhs` , `blvhs` , `navidnf` , `jrvhs` , `jlvhs` , `arvhs`)
- CD-ROM (`cagcdr` , `jccfrom` , `calncdr` , `nacdr` , `cacdr` , `arcdrom` , `accdrom` , `cacdrr`)
- Tape (`nabot` , `arcas` , `cccas` , `cybot` , `jccas` , `jccas` , `nacasnf` , `cacasnf` , `accas` , `cabot`)
- Microfilm (`ahmfc` , `camfr` , `armfm` , `cagmf` , `cadocmf` , `ahmfm` , `drmfmnp` , `drmfc`)

- CD (`arcd`, `nacd`, `cacdnf`, `jrcd`, `accd`, `cacd`, `cabocd`, `cccd`, `calncd`, `nabocdc`, `ncbocd`, `nybocd`, `jccd`, `nccd`)
- DVD (`cadvdnf`, `ncdvdnf`, `nalndvd`, `jrdvd`, `nadvd`, `ardvd`, `ccdvdnf`, `ncdvd`, `nadvdnf`)
- Vinyl Records (`carec`, `narec`, `jcrec`)
- Laser Disc (`nalas`, `nalasr`)

Now we have the itemType code for a specific storage device, we may find out their check out times per year, and use it as the index of popularity. The query is a little bit long, but the main idea is simple, check number of checkouts and categorize them using itemType code.

```
select *
from (
        select
            count(*),
            year(cout) as yr,
            CASE
                WHEN itemtype in (
                    'ucflpdr',
                    'naflpdr',
                    'acdisk',
                    'jrdisk',
                    'ardisk'
                ) THEN 'FloppyDisk'
                WHEN itemtype in (
                    'bcvhs',
                    'jcvhs',
                    'navid',
                    'alvhs',
                    'blvhs',
                    'navidnf',
                    'jrvhs',
                    'jlvhs',
                    'arvhs'
                ) THEN 'VHS'
                WHEN itemtype in (
                    'cagcdr',
                    'jccfrom',
                    'calncdr',
                    'nacdr',
                    'cacdr',
                    'arcdrom',
                    'accdrom',
                    'cacdrr'
                ) THEN 'CD-ROM'
```

```
                WHEN itemtype in (
                    'nabot',
                    'arcas',
                    'cccas',
                    'cybot',
                    'jccas',
                    'jccas',
                    'nacasnf',
                    'cacasnf',
                    'accas',
                    'cabot'
                ) THEN 'Tape'
                WHEN itemtype in (
                    'ahmfc',
                    'camfr',
                    'armfm',
                    'cagmf',
                    'cadocmf',
                    'ahmfm',
                    'drmfmnp',
                    'drmfc'
                ) THEN 'Microfilm'
                WHEN itemtype in (
                    'arcd',
                    'nacd',
                    'cacdnf',
                    'jrcd',
                    'accd',
                    'cacd',
                    'cabocd',
                    'cccd',
                    'calncd',
                    'nabocdc',
                    'ncbocd',
                    'nybocd',
                    'jccd',
                    'nccd'
                ) THEN 'CD'
                WHEN itemtype in (
                    'cadvdnf',
                    'ncdvdnf',
                    'nalndvd',
                    'jrdvd',
                    'nadvd',
                    'ardvd',
                    'ccdvdnf',
                    'ncdvd',
                    'nadvdnf'
                ) THEN 'DVD'
```

```
                WHEN itemtype in ('carec', 'narec', 'jcrec') THEN 'Vinyl Records'
                WHEN itemtype in ('nalas', 'nalasr') THEN 'Laser Disc'
                ELSE 'Other types'
            END AS StorageDevices
        from spl_2016.inraw
        group by
            yr,
            StorageDevices
    ) as res
where
    res.StorageDevices <> 'Other types';
```

We get the result:

https://drive.google.com/file/d/1L37SjSNTJkptmvvjg0in_uvT2Pt_B6do/view?usp=sharing

Findings:

65154	2005	CD
184	2005	CD-ROM
4	2005	FloppyDisk
1	2005	Microfilm
11306	2005	Tape
7540	2005	VHS
1337105	2006	CD
1151	2006	CD-ROM
1	2006	DVD
6	2006	FloppyDisk
8	2006	Microfilm
99942	2006	Tape
104314	2006	VHS
5	2006	Vinyl Records

In year 2005 and 2006, CD numbers has a huge boost, from 65154 to 1337105. This is probably because the data collection starts from 2005. The number of Tape and VHS also gained a huge boost (10 times). And there's still around 1k entries of CD-ROM.

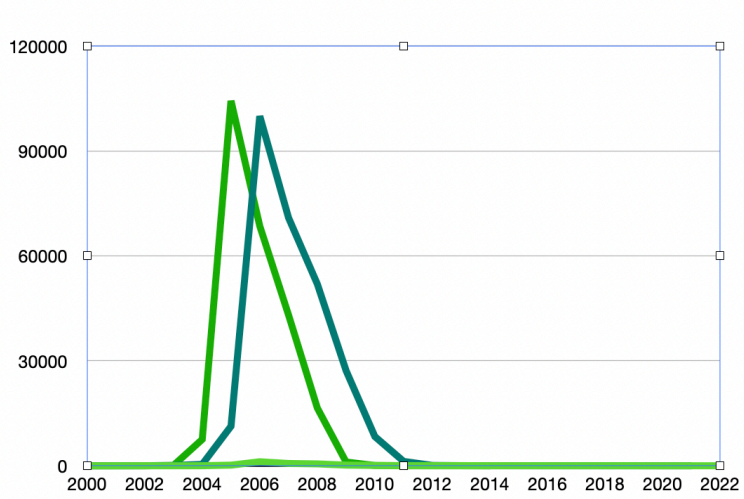
However, the number of Tape of VHS steadily decrease in the following years while CD still increases.

653	2007	CD-ROM
2	2007	FloppyDisk
17	2007	Microfilm
70871	2007	Tape
68502	2007	VHS
2	2007	Vinyl Records
2070298	2008	CD
539	2008	CD-ROM
3	2008	DVD
9	2008	FloppyDisk
7	2008	Microfilm
51936	2008	Tape
43076	2008	VHS
2	2008	Vinyl Records

At the year 2011, there's only 85 entries of VHS. And in 2012, there's only 110 entries of tape, while CD is around 1M.

1362713	2011	CD
52	2011	CD-ROM
52	2011	FloppyDisk
12	2011	Microfilm
1244	2011	Tape
85	2011	VHS
784698	2012	CD
28	2012	CD-ROM
21	2012	FloppyDisk
3	2012	Microfilm
110	2012	Tape
13	2012	VHS
799217	2013	CD
43	2013	CD-ROM

Visualization 01: VHS, Tape,CD-Rom by Year

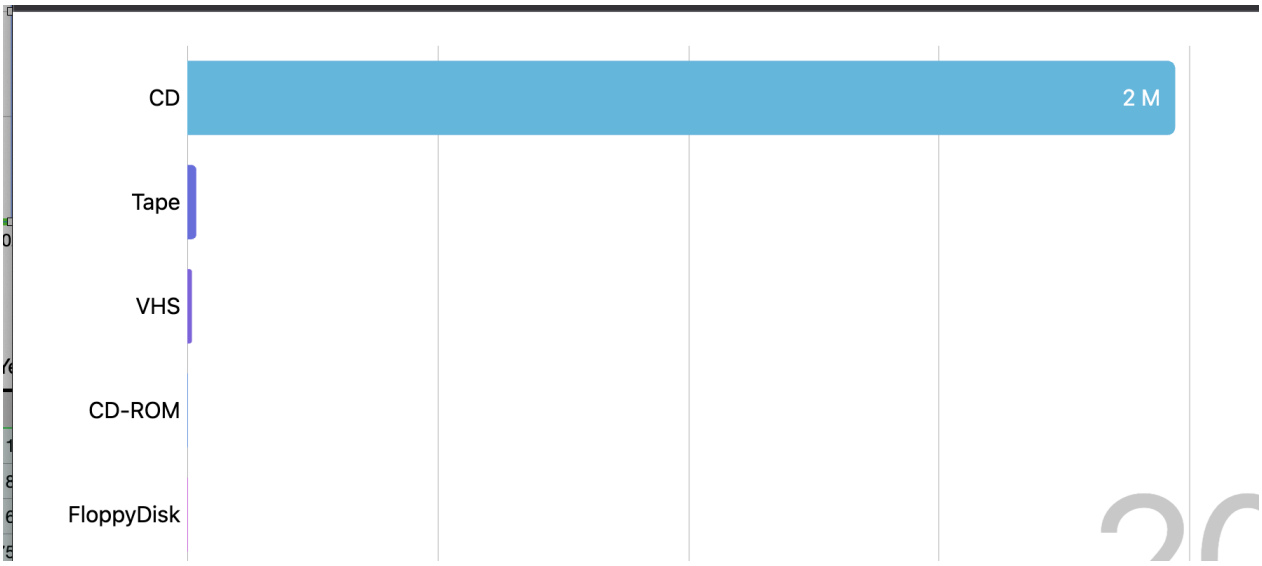


Tape and VHS follows the same pattern, they peaked around 2006, and sharply decreases in the next 5 years. As of today, there are only less than 10 entries per year.

Visualization 02: Storage Race

A dynamic way to visualize how the number of entries vary through years.

Referencing Demo from amChart5: https://naoyuki.top/DataViz/race/



The Tape/VHS to CD ratio is astonishingly high round 2006, and since then, CD entries has been almost 40 times more than Tape/VHS entries.

As of today, CD has 130k entries, 6.7% of its peak.

Tape has 1 entry, VHS has 2 entries. Tape entries has been negligible since 2012 and so is for VHS since 2010.

Possible Reason for 2006 peak:

- Because we start to collect data since late 2005, it's likely that data before 2005 were not fully documented. Thus, previous data can be inadequate. Since data after year 2005 is complete, there is a peak in 2006.

Possible Reason for other devices' unpopularity:

- As for Vinyl Records and DVD, there were only less than 10 entries each year. This is likely to be the issue of itemType.
- As for FloppyDisk and Microfilm, the data seems plausible. They are medium of several decades of history and have been discontinueted years ago.

What do people borrow within these category today?

Links to csv files of below queries: Link

For microfilm, use query below to check recent entries.

```
select *
from spl_2016.inraw
where
    itemtype in (
        'ahmfc',
        'camfr',
        'armfm',
        'cagmf',
        'cadocmf',
        'ahmfm',
        'drmfmnp',
        'drmfc'
    )
    and cout > '2014-01-01'
```

cout	cin	collcode	itemtype	barcode	title
2014-01-13 16:03:00	2014-02-19 13:17:00	cagmf	armfm	0000102964996	Troutman families of Kentucky
2014-03-21 11:55:00	2014-04-07 15:10:00	cagmf	armfm	0000102964434	Nicholas Mundy and descendants who settled in New Jersey in 1665
2014-04-22 12:56:00	2014-06-30 15:40:00	cagmf	armfm	0010042121110	records of the Olmsted Associates
2014-04-22 12:56:00	2014-06-30 15:41:00	cagmf	armfm	0010042121128	records of the Olmsted Associates
2014-08-15 11:30:00	2014-10-06 12:45:00	cagmf	armfm	0000102964459	Nash family or Records of the descendants of Thomas Nash of New Haven Conn
2015-02-20 10:20:00	2015-04-14 14:36:00	cagmf	armfm	0000102964004	History and genealogy of Davison Davisson Davidson families Arthur Henry Davison

They are mainly documentaries and historical motion pictures.

Same query is performed on VHS

```
select *
from spl_2016.inraw
where
    itemtype in (
        'bcvhs',
        'jcvhs',
        'navid',
        'alvhs',
        'blvhs',
        'navidnf',
        'jrvhs',
        'jlvhs',
        'arvhs'
    )
    and cout > '2012-01-01';
```

0:00	2012-01-27 12:45:00	naaavid	alvhs	001004062961	century of black cinema	VHS 791.43 C33394 1	79
7:00	2012-02-05 16:07:00	caread	bcvhs	0010037448007	GED connection Mathematics	VHS 373.12 G267 200	37
8:00	2012-05-05 14:09:00	naesli	bcvhs	0010045064101	American English I	VHS 428 AMERICA	42
4:00	2012-05-29 16:36:00	naesli	bcvhs	0010045064101	American English I	VHS 428 AMERICA	42
5:00	2012-10-10 14:13:00	naesli	bcvhs	0010045064101	American English I	VHS 428 AMERICA	42
3:00	2012-11-03 10:24:00	caread	bcvhs	0010037448007	GED connection Mathematics	VHS 373.12 G267 200	37
4:00	2012-12-09 15:55:00	naesli	bcvhs	0010045064101	American English I	VHS 428 AMERICA	42
6:00	2012-12-16 12:03:00	caread	bcvhs	0010037448007	GED connection Mathematics	VHS 373.12 G267 200	37
1:00	2012-12-20 13:36:00	naeslb	bcvhs	0010037597118	English for beginners Spanish	VHS 468.34 ENGLISH	46
1:00	2012-12-20 13:36:00	naeslb	bcvhs	0010038203237	English for beginners Russian	RUSSIAN VHS491.78 E	49
4:00	2013-01-04 08:25:00	caread	bcvhs	0010037448007	GED connection Mathematics	VHS 373.12 G267 200	37

Supringsly, the main titles people are looking for today in VHS format is about studying. There are mostly language materials. To check how many checkout within each title, I uses query below.

```
select count(title), title
from(
        select *
        from spl_2016.inraw
        where
            itemtype in (
                'bcvhs',
                'jcvhs',
                'navid',
                'alvhs',
                'blvhs',
                'navidnf',
```

```
                'jrvhs',
                'jlvhs',
                'arvhs'
            )
            and cout > '2012-01-01'
    ) as res
group by title;
```

	count(title)	title varchar(255)
1	20	American English I
2	1	century of black cinema
3	1	English for beginners Russian
4	1	English for beginners Spanish
5	6	GED connection Mathematics

Except for one title, the rest of them serve the purpose of learning. After typing `educational VHS` on eBay, there were more than 1000 items in this entry. It is safe to say VHS Tapes, till today, serve a educational purpose.

For CD-ROM

```
select *
from spl_2016.inraw
where
    itemtype in (
        'cagcdr',
        'jccfrom',
        'calncdr',
        'nacdr',
        'cacdr',
        'arcdrom',
        'accdrom',
        'cacdrr'
    )
    and cout > '2021-01-01';
```

1	3	diaries of Delridge
2	3	diaries of West Seattle
3	1	Vaccination a history of vaccines
4	1	Comprehensive management and use plan final environm
5	1	wildlife investigator series
6	1	Earthquakes and eruptions
7	1	Atlas of United States mortality
8	1	Using and presenting digital evidence in the courtroom n

CD-Rom titles seem more diverse, there are titles about history, courtroom, geography and advanced technology. Most of them are still within Geography category.

For Tape

```
select *
from spl_2016.inraw
where
    itemtype in (
        'nabot',
        'arcas',
        'cccas',
        'cybot',
        'jccas',
        'jccas',
        'nacasnf',
        'cacasnf',
        'accas',
        'cabot'
    )
```

```
16      and cout > '2016-01-01'
```

count(title)	title varchar(255)
16	Cant take me home
5	Twinkle twinkle little star
4	Llaman a la puerta
3	history of Native America
1	Selected works for lute
1	I miss my Swiss My Swiss miss misses me walk around

It is no surprise that most of these checkouts are within music category. One excepetion is about the history of Native America.

Visualization:

To make it easier to understand what people are looking for these days, I made a word cloud visualization on the title within these categories. The most popular entries are, American English, History and Seattle.

Conclusions:

- Before the year 2006 and 2007, there were many VHS and Tapes and CD-Roms on the market. However, as streaming services became popular and new advanced storage devices released to the market, these legacy devices is subject to a sharp decrease in their popularity. From 10k checkouts per year to less than 10 checkouts each year. While CD dominated the market for more than a decade. As of today, CD has 130k entries, 6.7% of its peak.
- These items today mainly serve educational purposes, mainly language learning, geography and history knowledge. It is nice to see that there are still many VHS tapes on the market, and most of them serve the purpose of language learning, which is surprising.

Reference:

- SPL OpenData: https://data.seattle.gov/w/pbt3-ytbc/2myu-6xk5?cur=tlm8xU9wwlh
- amChart Demo: https://www.amcharts.com/demos/
- eCharts wordCloud: https://github.com/ecomfe/echarts-wordcloud